COMPARISON OF MODELLING BETWEEN ARX AND OUTPUT ERROR MODEL

MOHAMAD HAIDIR BIN NASIR

A report submitted
in fulfilment of the requirements for the degree of
Bachelor of Mechanical Engineering with Honours

Faculty of Mechanical Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2018

DECLARATION

I declare that this thesis entitled "Comparison Of Modelling Between ARX And Output-Error Model " is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ....

Name : MOHAMAD HAIDIR BIN NASIR

Date : ....

DEDICATION

To my beloved mother and father

APPROVAL

I hereby declared that I have read this report and in my suggestion this report is sufficient in terms of scope and quality for the award of the degree of Bachelor of Mechanical Engineering (BMCG).

Signature :…………………………………………………….

Supervisor Name : DR. MD. FAHMI BIN ABD. SAMAD

Date :…………………………………………………….

ABSTRACT

This research presents the comparison of modeling between ARX and Output-Error model in System Identification. System identification is an idea of simulating the real event to develop a mathematical modeling. The purposes of this research are to simulate modelling using ARX and Output-Error models and to compare the modelling performance of ARX and Output-Error models based on selected indicators. Decidedly, the performance indicators used in this research are best fit, Final Prediction Error (FPE), and Mean Square Error (MSE). MATLAB R2015b software provides "ident", which is a command of GUI environment that already set in System Identification Toolbox, which is running on preprocessing and system identification operations. Therefore, all the data have been generated in the System Identification Toolbox along the research studies. Least square method was used to find the line of best fit for a data set. Generally, the analysis that have been done showed the structure of Output-Error model is quite better than ARX model due to its largest best fit value and also smallest FPE and MSE values. Moreover, a real data that has been implemented also showed that Output-Error model is the best performance. In conclusion, the result of the graph in the analysis part were influenced by the data distribution.

ABSTRAK

Kajian ini membentangkan perbandingan pemodelan antara model ARX dan Ralat Output dengan menggunakan Pengenalpastian Sistem. Pengenalpastian sistem adalah idea mensimulasikan peristiwa sebenar untuk membangunkan pemodelan matematik. Tujuan penyelidikan ini adalah untuk mensimulasikan pemodelan menggunakan model ARX dan Output-Error dan untuk membandingkan prestasi pemodelan model ARX dan Output-Error berdasarkan petunjuk terpilih. Penunjuk prestasi yang digunakan dalam penyelidikan ini adalah kesesuaian, Ralat Ramalan Akhir (FPE), dan Ralat Purata Kuasa Dua (MSE). Perisian MATLAB R2015b menyediakan "ident", yang merupakan perintah persekitaran GUI yang telah ditetapkan dalam Kotak Pengenalpastian Sistem, yang dijalankan pada operasi pengolahan dan sistem pengenalan awal. Oleh itu, semua data telah dihasilkan dalam Kotak Pengenalpastian Sistem sepanjang kajian penyelidikan. Kaedah kuasa dua terkecil digunakan untuk mencari garis yang paling sesuai untuk satu set data. Secara amnya, analisis yang telah dilakukan menunjukkan struktur model Ralat Output adalah agak lebih baik daripada model ARX kerana nilai kesesuaian terbaiknya dan juga nilai FPE dan MSE terkecil. Selain itu, data sebenar yang telah dilaksanakan juga menunjukkan bahawa model Ralat Output adalah prestasi terbaik. Kesimpulannya, hasil graf dalam bahagian analisis dipengaruhi oleh pengagihan data.

ACKNOWLEDGEMENTS

It is a really great opportunities for students especially the final year students of the Mechanical engineering course to experience the Final Year Project (FYP) process. Due to that, I am very thankful and felt very lucky to be part of it. I am also looking forward to work along with the great and exciting people in order to complete the project. I am realised that it is not easy to get the opportunities to work with the great people.

First and foremost, I would like to take this opportunity to express my sincere acknowledgement to my supervisor Dr. Md. Fahmi Bin Abd. Samad from the Faculty of Mechanical Engineering Universiti Teknikal Malaysia Melaka (UTeM) for his essential supervision, support and encouragement towards the completion of this thesis.

Particularly, I would also like to express my deepest gratitude to my FYP partners, Mr. Che Izzat and Miss Nadhirah for helping me in completing this research. Thanks for being there when I need some help and not understand some parts of the research.

Special thanks to all my peers, my late mother, beloved father and siblings for their moral support in completing this degree. Lastly, thank you to everyone who had been to the crucial parts of realization of this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ABBREVIATIONS		TITLE
BJ	-	Box-Jenkins
ARX	-	Autoregressive wih exogenous
OE	-	Output-Error
SISO	-	single input single output
GUI	-	Graphical User Interface
FRD	-	frequency response data
FPE	-	Final Prediction Error
MSE	-	Mean Square Error

CHAPTER 1

INTRODUCTION

1.1 Background

In our environment, people learn to control their actions by predicting their effect. For example, if one pushes a ball downhill, it rolls and then we are able to deal with much more complicated challenges like walking, biking, swimming and running. A coherent model is used to describe the aspects that we are interested in and do not try to use it to describe all of reality. All of his reality is split up and efforts are concentrated on just one part of reality at a time and this is called a system. The rest of nature represent the environment of the system. Interaction between the system and its environment are described by input and output ports. In our reality, we are always make errors when measuring a length, weighing a mass, the current or voltage. This is because the instruments we use are not perfect and the models are also imperfect. Many systems are not deterministic and it is impossible to predict exactly their output as this show a stochastic behaviour. The model then split up into a deterministic part and a stochastic part. The deterministic aspects are represented by the mathematical model while the stochastic behaviour is modelled as a noise distortion.

Identification is a good method for extract accurate models of complex systems from noisy data. The aim of identification theory is to provide systematic approach to fit the mathematical model to the deterministic part and reduce the noise distortions as much as possible. There are many types of model used in system identification such as ARX model, ARMAX model, Output-Error model, and Box-Jenkins model (BJ model).

In this research, the comparison of modelling between ARX and Output-Error models will be studied. These two models have different general models as the advantages and disadvantages between the two also will be investigated in this project. The project shall be done by using a software called MATLAB.

1.2 Problem Statement

System identification is a method to identify important variables as it is crucial in industrial needs. This situation poses an issue of which model is suitable for identification process. There are many general models such as ARX model, ARMAX model, Output-Error(OE) model and Box-Jenkins model (BJ model). Therefore, this project will try to clarify the difference between ARX model and Output-Error model.

1.3 Objectives

The objective of this study is :

1. To simulate modelling using ARX and Output-Error models
2. To compare the modelling performance of ARX and Output-Error models based on selected indicators.

1.4 Scope Of Project

There are a few scopes as the limitation that need to be followed in this project.

1. All simulation will be done using the identification graphical user interface (GUI) in MATLAB.
2. Data acquisition will be made based on simulated data in the form of single input single output (SISO). This system will also be made by using MATLAB.
3. The performance of modelling will be judged based on several selected indicators provided in the GUI.
4. The parameter estimation stage will be done by using least square method

1.5 General Methodology

In order to complete this project and to achieve the objectives, there are certain methods that need to be followed as shown in Figure 1. More details about methodology will be provided in Chapter 3.

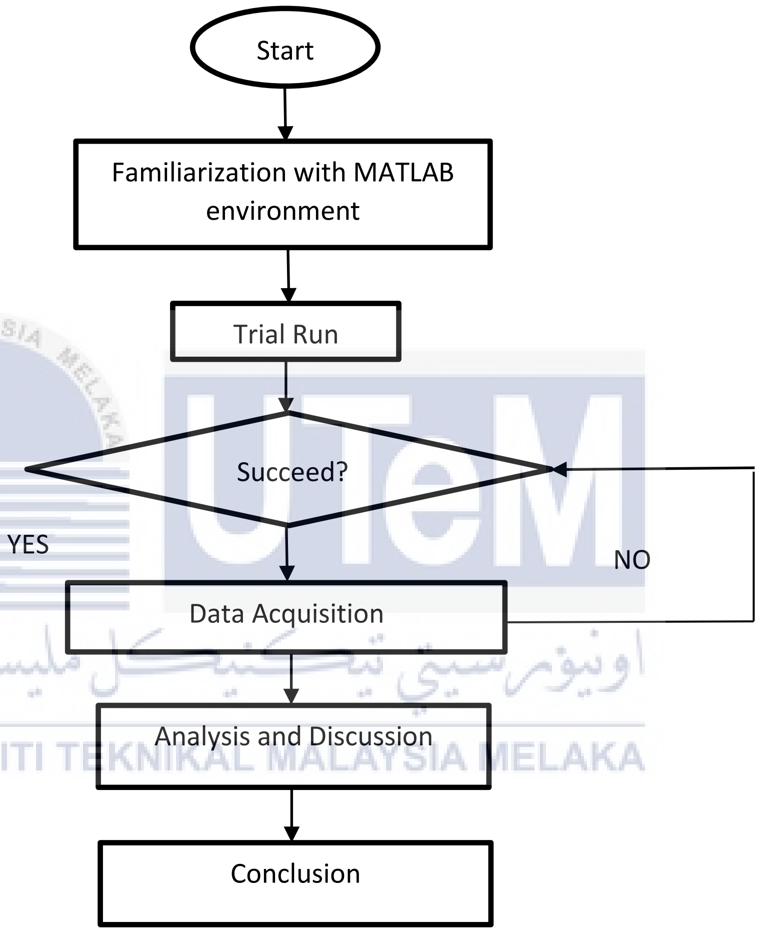


Figure 1: Flow chart of Methodology

CHAPTER 2

LITERATURE REVIEW

2.1 System Identification

2.1.1 Introduction

System identification is an idea of simulating the real event to develop mathematical modeling. This system contains experimental planning, selection of model structure, criteria, parameter estimation and model validation (Astrom and Writtenmark, 1997). The availability of appropriate models for process dynamics is essential to analyze and design a control system. By using frequency response, one could get the transfer function of a process. In fact, it contributes to the success of classical control theory in Laplace transform. A digital control was developed together with the system identification methods.

2.2 Basic Steps Of Identification Process

There are some basic steps in identification process as below:

2.2.1 Gather Information About The System

Basically, if we want to build a model for a system, then we need to find and collect information about it. There are some ways to get a good information which are done by watching the natural fluctuations and experiment that could be more efficient (Pintelon, 2012).

2.2.2 Select A Model Structure To Represent The System

According to Pintelon and Schoukens (2012), there are many possibilities of mathematical models that can be used to represent a system such as:

- Parametric versus nonparametric models

The comparison between parametric model and nonparametric model is their characteristics. A limited number of quantities known as parameters of model is used to describe the system of parametric model. Meanwhile, for nonparametric model, a system function is measured at a huge number of points to be characterized. In this condition, a transfer function of a filter is used as an example. The piston's equation of motion and poles and zeros are the parametric model while the filter described by its impulse response at a huge number of points is the nonparametric model.

- White box models versus black box models

A construction of a model needs physical laws based on experimenter's insight and skills such as Kirchoff's laws and Newton's laws. Taking a modeling of a loud speaker as an example, needs wide understanding of electrical, mechanical and acoustical phenomena. The result may lead to a physical model based on detail knowledge of the internal function system. That model is known as white box model. Another approach is black box model where its mathematical model is measured by input and output relation. White box model is much better compared to black box as it applies knowledge into working principles of a system. But, a black box model might be acceptable if it only in purpose of predicting the output.

- Linear models versus nonlinear models

Most of the nonlinear system theory is relative to the linear system by assuming the behavior can be linearized in the operation region. For example, the distortion of an amplifier is described by a nonlinear model. In other case, if the linear behavior is found dominant and is the only interest then a linear model might be sufficient to represent the transfer characteristics.

- Linear-in-the-parameters versus nonlinear-in-the-parameters

When parameters and error that is minimized produced a linear relation then the model is known as linear-in-the-parameters. For example, $\varepsilon = y - (a_1 u + a_2 u^2)$ shows linear parameters, a_1 and a_2, but it actually describes a nonlinear system. But, the following equation below has nonlinear parameters, b_0 and b_1 but it describes a linear system.

$$\varepsilon(j\omega) = Y(j\omega) - \frac{a_0 + a_1(j\omega)}{b_0 + b_1(j\omega)} U(j\omega)$$

2.2.3 Match the Selected Model Structure to the Measurements

The model structure gained should be matched with the available information about the system. Minimizing a criterion that measures an advantage of the fit is the best way. The selection of this criterion is essential part as it determines the stochastic properties of the final estimator.

2.2.4 Validation of Selected Model

This is the final basic step of identification process. Its aim is to find out that the selected model describes the data availability or any left indications not well modeled from the data. It is not always for the model with less error can be accepted. But, the preferred model is a simple model that describes the system within user-specified error bounds. The user will be guided by the provided tools in this process. The remaining error will be separated into classes respectively. For example, the classes may be nonlinear distortions and unmodeled linear dynamic.

Another researcher, Astrom and Wittenmark (1997) in the same field, provide steps that have similarities with the steps of identification process from above. One of them is experimental planning which is gaining information first about the methods to be used. Second is the selection of model structure derived from prior knowledge of the process and disturbance. The prior knowledge can be described as a linear system. Then, general representations or called black-box models are used for that linear system.

$$A(q)y(k) = B(q)u(k) + C(q)e(k)$$

A representative example is the different-equation model where u is the input, y is the output, and e is a white-noise disturbance. Meanwhile, A, B, and C are considered as unknown parameters. The physical law sometimes is possible to be applied in order to derive models of the process that only have some unknown parameters. Then, the model form may be as equation below.

$$\frac{dy}{dx} = f(x, u, v, \theta)$$

$$y = g(x, u, e, \theta)$$

Where θ is a vector of unknown parameters, x is the state of the system, and v and e are the disturbances. The other steps are criterion selection, estimation of parameter, and model validation.

2.3 Discrete-Time System And Continuous-Time System

A computer-controlled system can be represented schematically as in Figure 2.1. The output of $y(t)$ is a continuous-time signal. Analog-to-digital (A-D) converter converts the output into digital form. The conversion is set at sampling time, t_k. The converted signal, $\{y(t_k)\}$ is interpreted as a sequence of numbers. Then, the algorithm processes the measurements and forms a new sequence of numbers, $\{u(t_k)\}$. Next, a digital-to-analog (D-A) will convert this sequence to an analog signal as continuous signal.

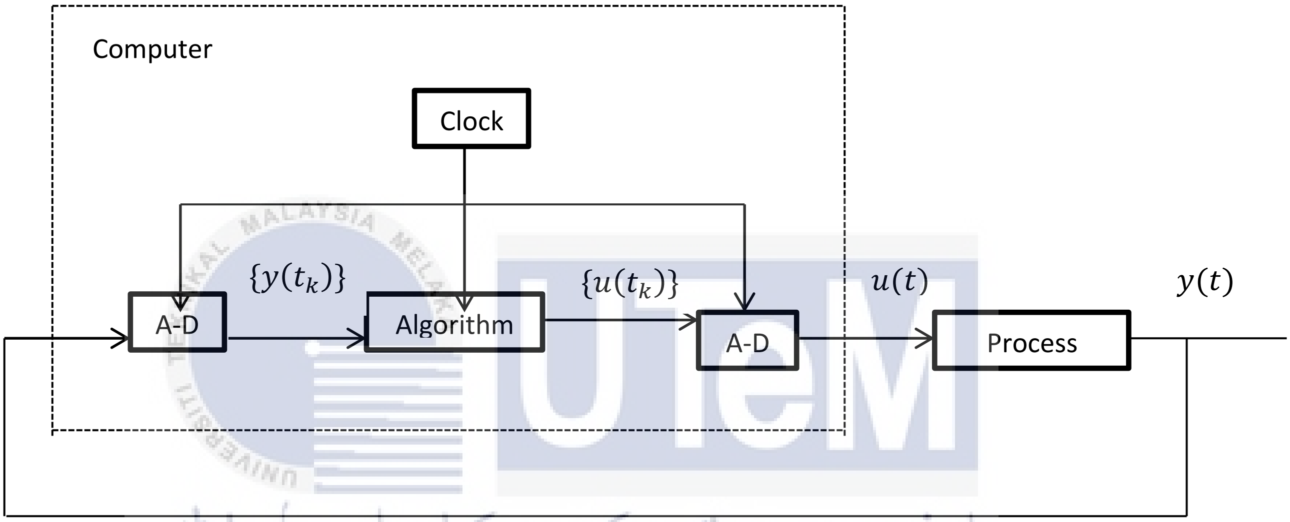


Figure 2.1: Schematic diagram of a computer-controlled system (Astrom and Wittenmark, 1997)

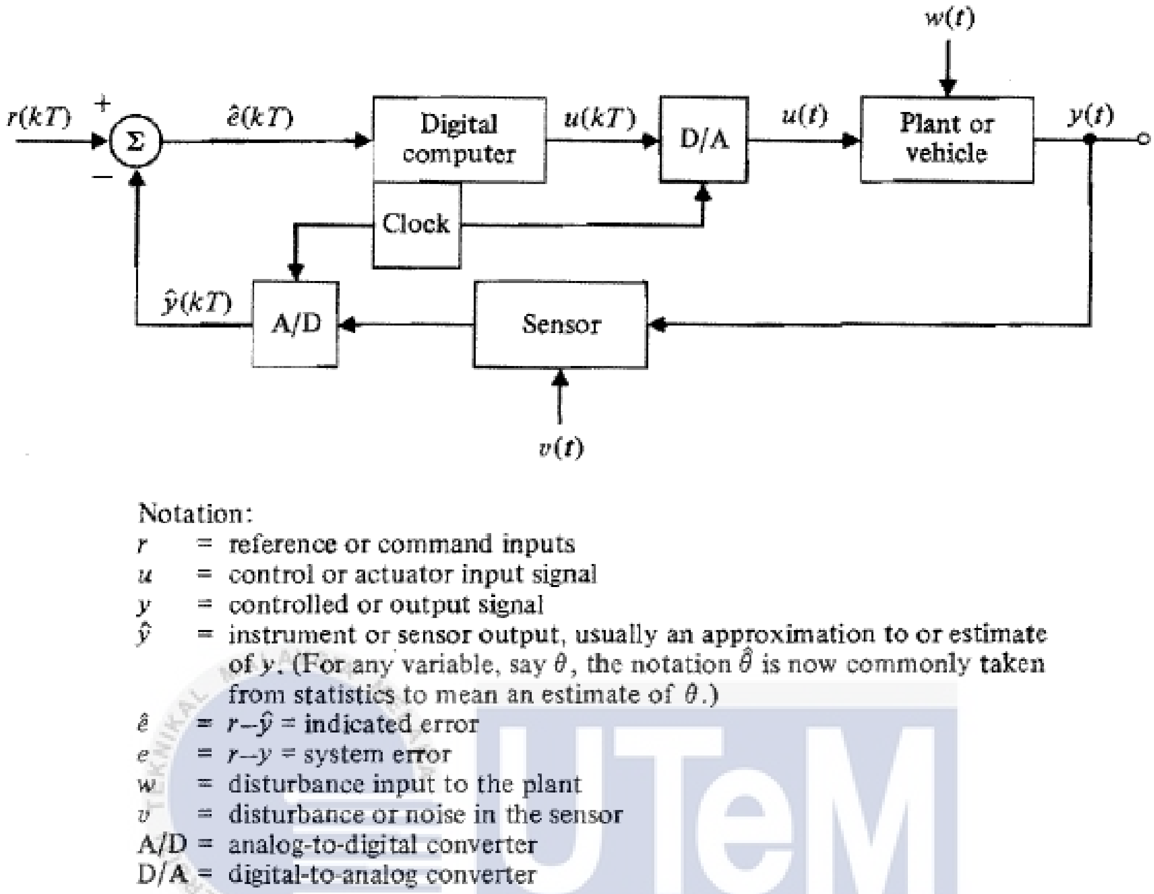


Figure 2.2: Block diagram of a basic control system

Figure 2.2 is taken from another research of basic control system (Franklin et al., 1998). The plant output, *y(t)* have to follow the reference input, *r(t)*, even where there are presences of disturbance, *w(t)*. and errors in the sensor *v(t).* In order to generate the control action, a digital device is used to perform tasks as in Figure 2.2. Analog-to-digital (A-D) converts a physical variable like an electrical voltage and interprets it into a stream of numbers. The A-D converter works on sensor output, $\hat{y}$, and send it as a sequence of numbers to digital computer despite of having an error, $\hat{e}$. All the numbers are set to reach the digital computer at a sampling time, T. The clock gives a pulse or interrupt at every T seconds and A-D converter forwards a number to digital computer every time interrupt appears.

2.4 ARX And ARMAX Processes

Linear systems accompanied by white noise can generate large classes of stochastic processes. Let $\{e(k),\ k = \ldots, -1, 0, 1, \ldots\}$ be discrete-time white noise. The process generated by

$$y(k) = e(k) + b_1 e(k-1) + \ldots + b_n e(k-n)$$

is known as moving average, MA process where *y* is the output of continuous-time signal. Then, the process generated by

$$y(k) + a_1 y(k-1) + \ldots + a_n y(k-n) = e(k)$$

is called an autoregression, AR process. When AR process with exogenous signal, X got the ARX equation as below

$$y(k) + a_1 y(k-1) + \ldots + a_n y(k-n) = b_0 u(k-d) + \ldots + b_m u(k-d-m)$$

Where *u* is the exogenous signal, X. The parameter *d* represents input time delay.

The process below is called an ARMA process.

$$y(k) + a_1 y(k-1) + \ldots + a_n y(k-n) = e(k) + b_1 e(k-1) + \ldots + b_n e(k-n)$$

The process below is called an ARMAX process which means an ARMA process with exogenous signal, X (Astrom and Wittenmark, 1997).

$$y(k) + a_1 y(k-1) + \ldots + a_n y(k-n) = b_0 u(k-d) + \ldots + b_m u(k-d-m) + e(k)$$
$$+ c_1 e(k-1) + \ldots + c_n e(k-n)$$

2.5 Output Error Model

The identification method of the output error model is the prediction error method.

$$y(t) = \frac{B(q)}{A(q)} u_0(t) + e(t) = G_0(q) u_0(t) + v(t)$$

It corresponds to white noise disturbances added to the output of the system (Schoukens, Pintelon, and Rolain, 2012).

2.5.1 Hammerstein Output-Error System

Hammerstein nonlinear output-error model is a linear dynamical block following a memoryless nonlinear block. The information vector in the identification model contains unknown variables. A common system industry known as Hammerstein and Wiener (H-W) is combination of linear time-invariant blocks and static nonlinear systems. There are two classes of identification approaches in H-W models. They are iterative and recursive. Recursive algorithm can be on-line implemented but iterative cannot.

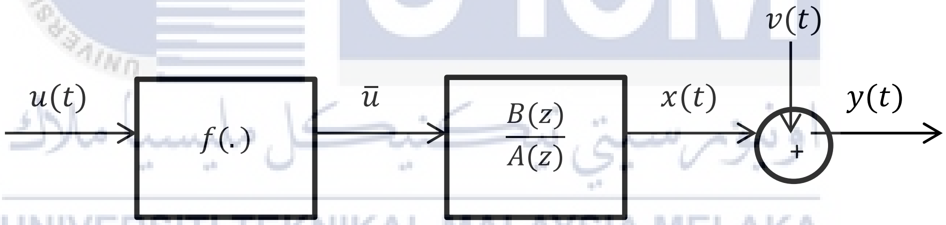


Figure 2.3: The Hammerstein nonlinear output-error system (Ding et al, 2007)

2.6 Least Square Method

A form of mathematical regression analysis or called least square method is used to find the line of best fit for a data set. The line of best fit is a straight line drawn through the center of a group of data points plotted on a scatter plot. Scatter plots describe the results of gathering data on two variables. The line of best fit identifies whether these two variables appear to be correlated and can be used to help identify trends occurring within the dataset (Staff, 2016). This method also used to identify parameters in dynamic system (Astrom and Wittenmark, 1997). A straight line is created by least square method to minimize the square errors generated by the related equations based on the model (Staff, 2016).

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter will explain the details about the whole approach of this study. Before going further, it is important to have a brief flow chart as a guidance of the progress and flow of this study. Furthermore, the presence of the flow chart can make the progress implemented be made orderly and flow of all the works involved become more organized and systematic. The flow chart consists of the initial process of identification by MATLAB R2015b Edition software simulation. The icon for the software is as shown in Figure 3.1. Meanwhile, Figure 3.2 shows the whole body of the flow chart for this study.

Figure 3.1: MATLAB R2015b Software

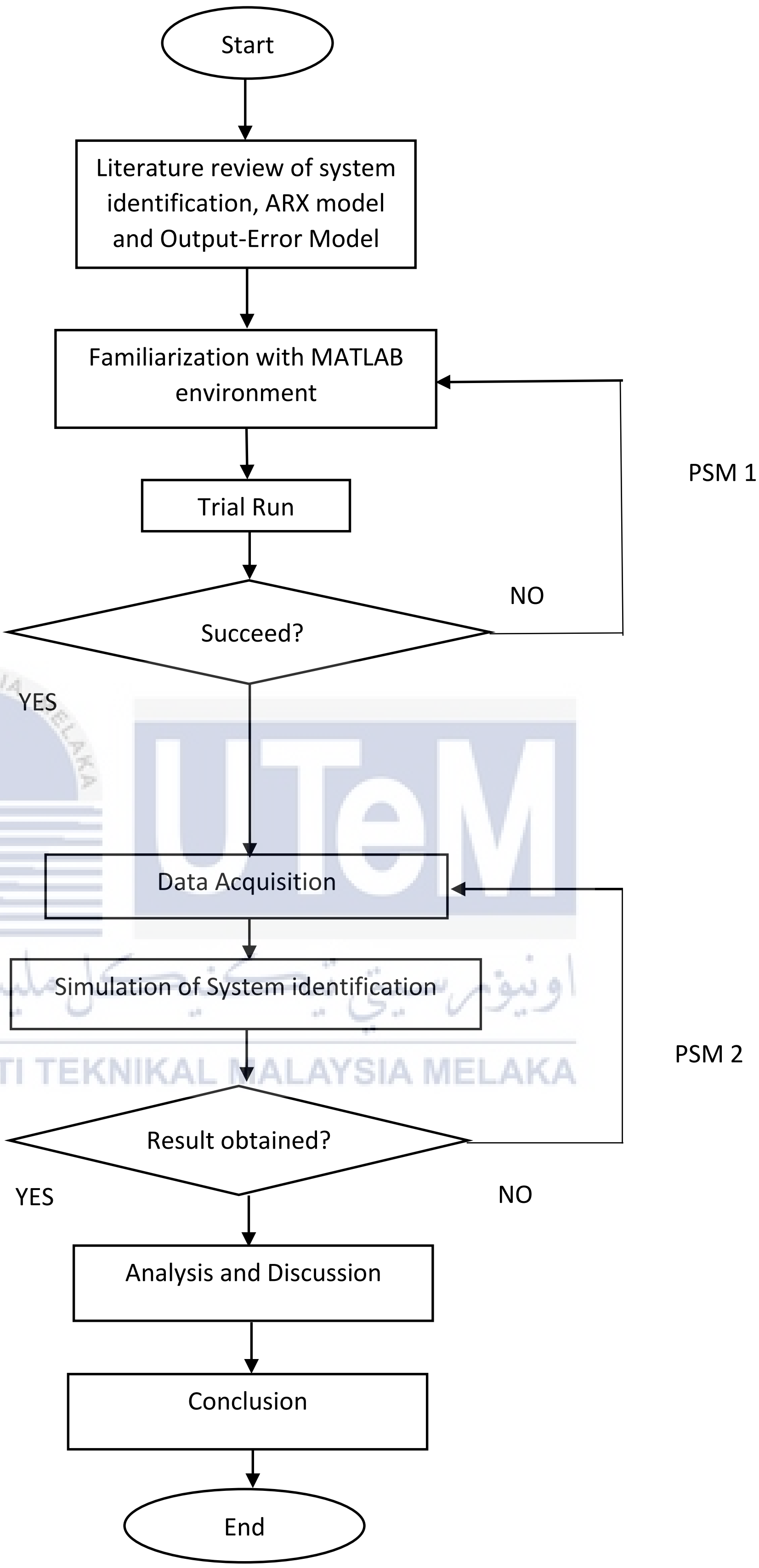


Figure 3.2: Flow chart of Methodology

3.2 Familiarization with MATLAB Environment

MATrix LABoratory or known as MATLAB is a powerful numerical analysis program and computation. Analyzing data, developing algorithms, and creating models and applications can be obtained by using this software (Pintelon and Schoukens,2012). By comparing other programming software such as C++ or Java, MATLAB can search for various approaches and get the solution faster. Besides, many applications like signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology can be explored by using it. Therefore, MATLAB, the language of technical computing is widely used by engineers and scientists in industrial area.

Next, mathematical models of dynamic systems that is not easily modeled from first principles or specifications can be built up from measured input-output data by using Simulink blocks, and MATLAB functions provided by System Identification Toolbox. Continuous-time and discrete-time transfer functions, process models, and state-space models can be determined by using time-domain and frequency-domain input-output data. The Algorithms for embedded online parameter estimation is also provided in this toolbox. The graphical user interface of System Identification Toolbox is as shown in Figure 3.3.

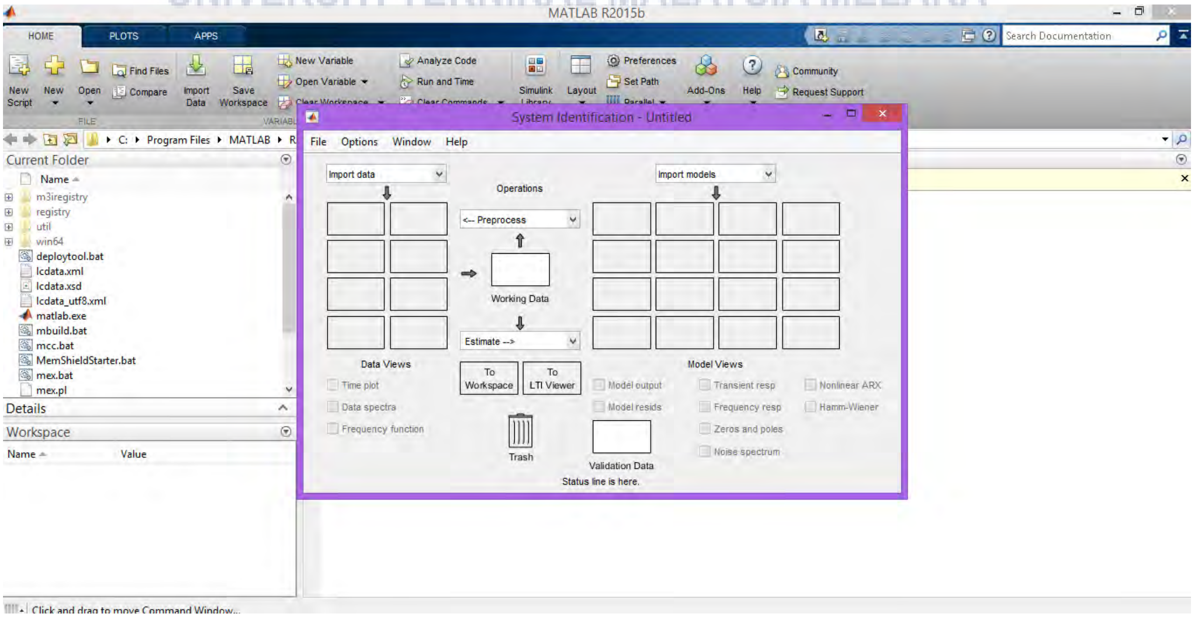


Figure 3.3: System Identification Toolbox

There are multiple approaches to system identification in MATLAB. Therefore, it is quite difficult especially for the beginners who are not familiar to apply MATLAB commands and system identification theory. Hence, a Graphical User Interface (GUI) is recommended. It provides the procedures of system identification. By using the GUI, the troubles faced by beginners in developing system identification might be minimized. Thus, "Ident" is the command to call upon the GUI environment that is already set in System Identification Toolbox., which is suitable for preprocessing and system identification operations. Some of the functions of the GUI in System Identification Toolbox are generating input signals, collecting and preprocessing input-output signals, executing the algorithm of system identification and designing control systems.

3.3 Trial Run

This section explains the steps on this trial run. The data taken provided in MATLAB software. It is used on the purpose as a preliminary run for System Identification and the functionality in GUI. To import data into the GUI:

1. “Load dryer2” is entered in the MATLAB command window as shown in the following figure. Data variables u2 and y2 appeares in the workspace. An actual hair dryer is presented by these data variables. The input (u2) is the heating power and the output (y2) is the temperature of the outflow air.

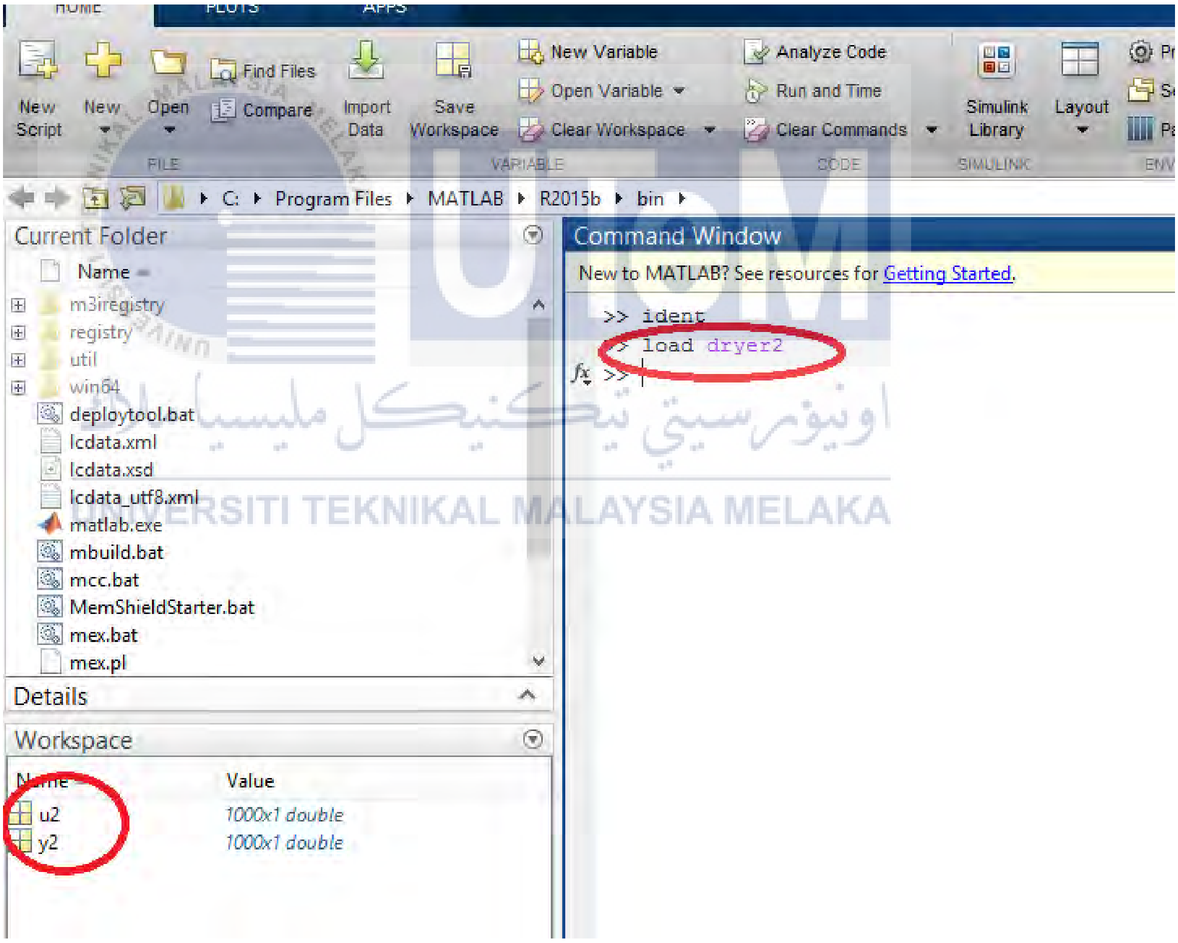


Figure 3.4: MATLAB command window

2. In Import data popup menu, the Time domain data is selected and this action opens the Import Data dialog box as shown in Figure 3.6.

Figure 3.5: Import data

3. In the Import Data dialog box, u2 is key in for Input and y2 for Output in the workspace variables.
4. In the Data Information, a Sample time: 0.08 and Starting time 1 are put.
5. Key in any variable name in the box stated Data name:. For example: dry
6. A dialog box will open when clicking on button More and a box marked Notes for any text wished to be entered. For example, “this is my first time using this software”. The sampling interval is the time between successive data samples in an experiment and must be the numerical time interval at which the data is sampled in any units. For example, insert 0.8 if the data was sampled every 0.8 s, and insert 1 if the data was sampled every 1 s. Modeling process usually used the sampling interval and start time to calculate the sampling time instants for time-domain data.

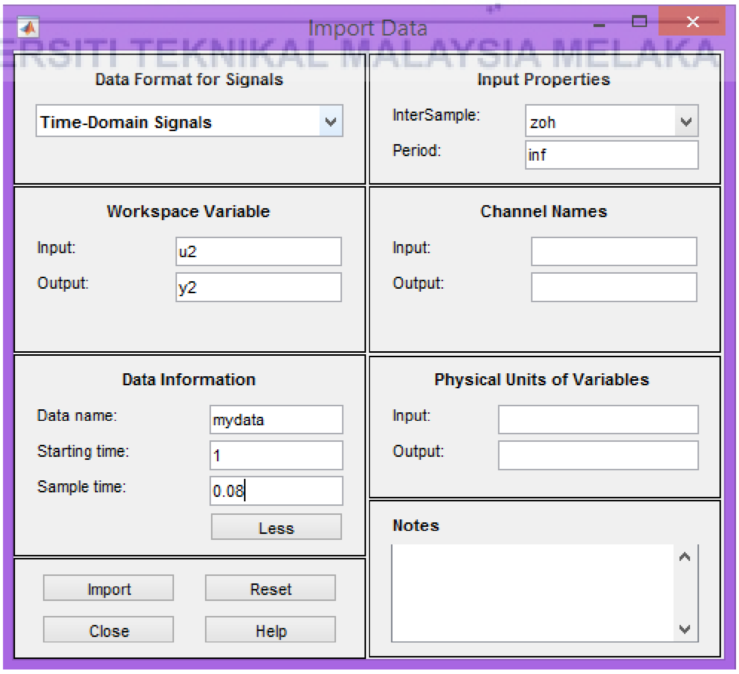


Figure 3.6: Import data dialog box

7. The data that have been key in after pressing the Import button is presented as Figure 3.7.

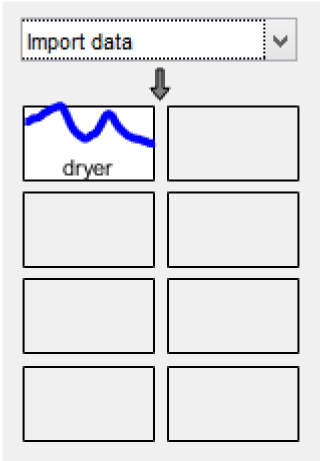


Figure 3.7: Set data

8. On Data Views, the Time Plot is selected to open a plot figure. Time plot is known as a function of time. The horizontal axis on the time plots figure resulted from the sampling interval during model estimation in step 4 and 6. Transforming a time-domain signal to a frequency-domain signal will generate a discrete Fourier transform (DFTs) developed by this sampling interval. The following figure shows the time plots of the set sampling interval.

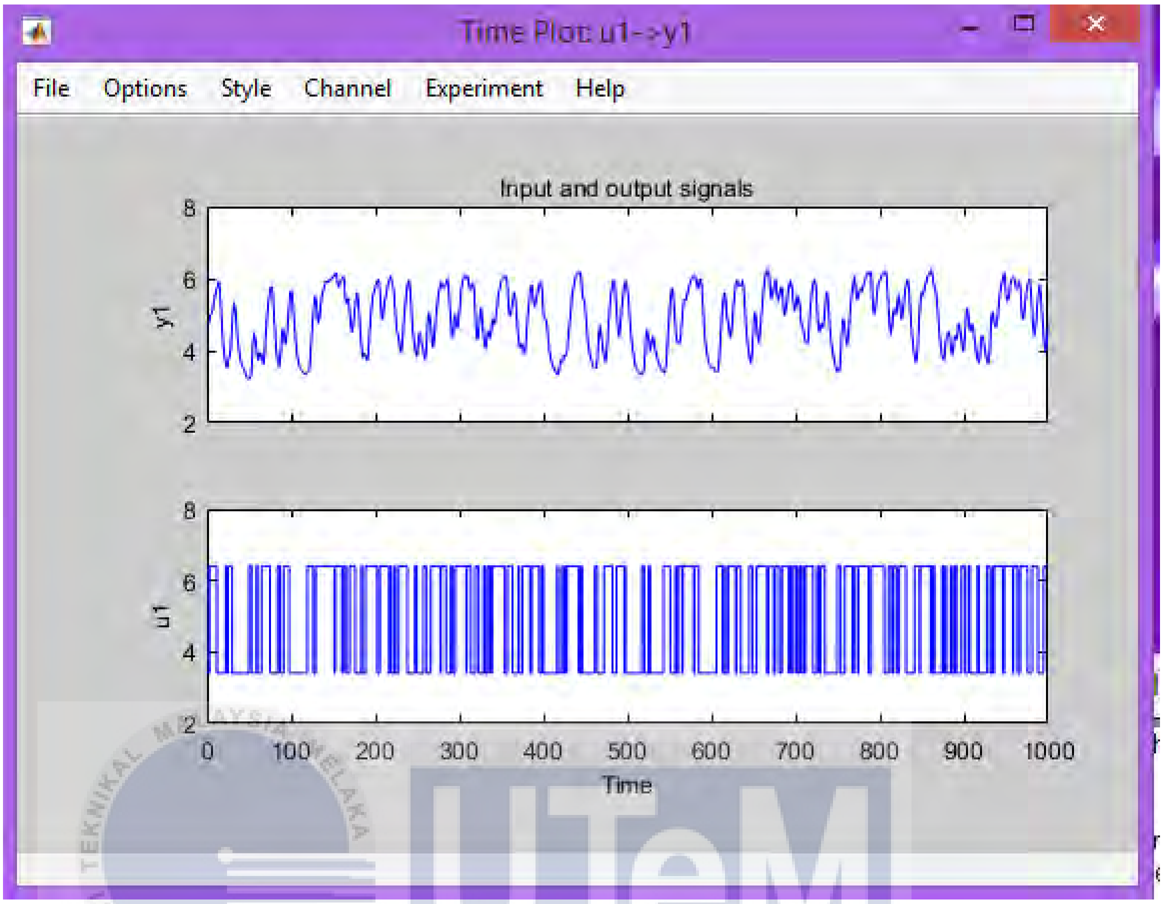


Figure 3.8: Time plot

9. The plot is studied by using multiple options from the plot character through the menu bar.
10. Taking an example from one of the option provided in menu bar which is Style-the Zoom was enabled. Next, rectangles can be formed by clicking the left mouse button in the plot. The area of that rectangle enlarges to fill the axes when releasing the button. Then double-click to back to the original plot. For more information about help, enter "help zoom" at the MATLAB command window.
11. To remove the constant levels in the data sequences, Remove means is selected from the Preprocess popup menu. A new data set will be inserted into the Data board.
12. The new data set can be seen automatically in the Time plot figure. To see the new plots, Autorange is selected under the Options menu. To unplot a data set, the icon is clicked in order to deselect it. Next, the new data set with a "d" attached at the end of its name is dragged into working data to explore more.

13. Double click on a data set icon and a new dialog box appeared containing information about the data selected.
14. Select range is chosen in the Preprocess popup menu and Figure 3.9 will be opened.

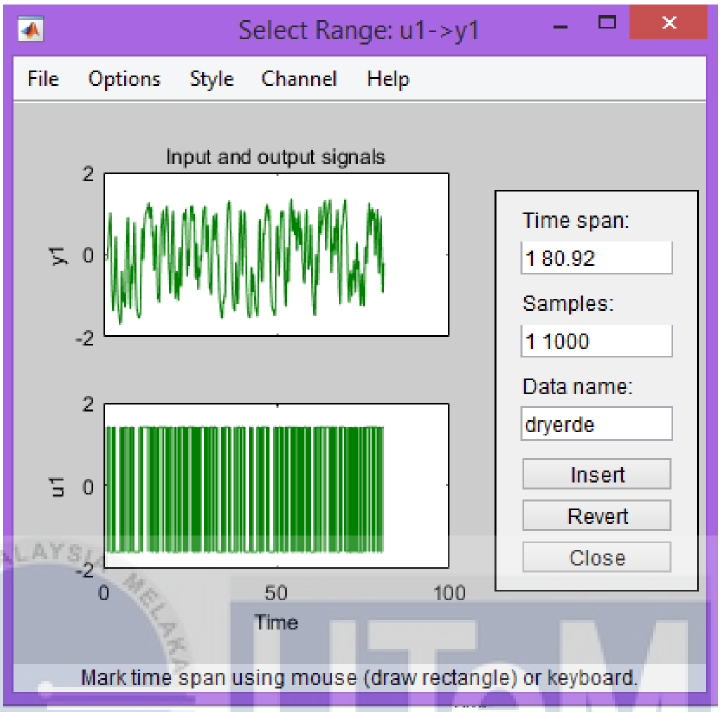


Figure 3.9: Select Range dialog box

15. The interval from 1 to 50 seconds is chosen to draw a rectangle by clicking the mouse button down to select a portion of the data to be used for estimation purposes. The rectangle can be erased and Revert button is pressed to restart.
16. Insert button is pressed and the selected data range is added as a new data set onto the Data board.
17. Return to the portion of the data set to allocate data be used for validation process. This is done by repeating step 15 by selecting the interval from 50 seconds to the end. There will be two data sets dryde and drydv on the data board as shown in following figure.

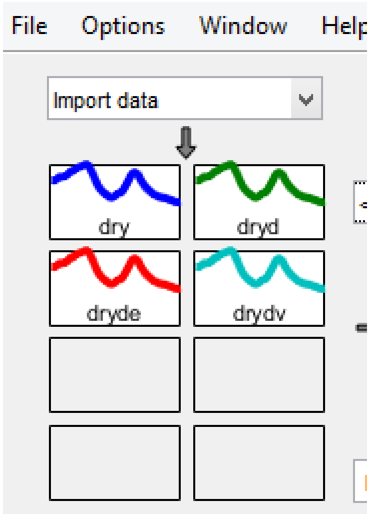


Figure 3.10: Set data

18. In the Data board of the application, a data set is dragged to Working Data icon for estimation. Meanwhile, a data set is dragged to the Validation Data icon for validation purposes.
19. As described in step 18, the third data set (dryde) is used for estimation purposes and the fourth data set (drydv) for validation purposes.
20. To estimate Finite Impulse Response (FIR), Correlation listed in the Estimate popup menu is selected. Figure 3.11 is opened by clicking Estimate. A model will be added to the Model board.

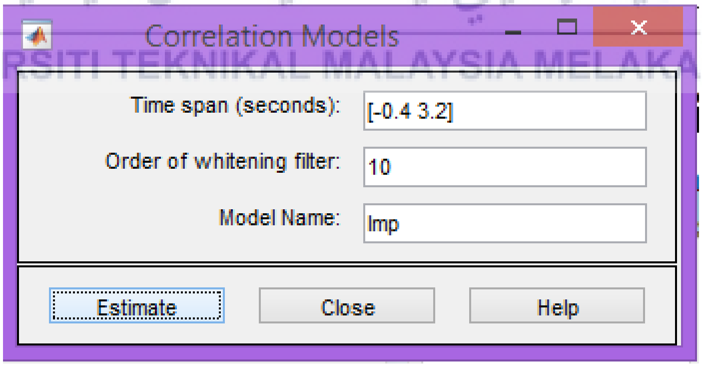


Figure 3.11: Correlation Models

21. Under Model Views, Transient resp as shown in the figure below is selected and a plot figure of Transient response appeared.

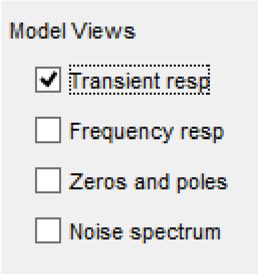


Figure 3.12: Model Views

22. To estimate the frequency response data of the system (FRD), Spectral Models is selected under Estimate popup menu. A dialog box will be appeared and the button Estimate is pressed to add a model to the Model board.
23. To view the result, the "Frequency resp' is clicked at the same figure shown in step 21. A figure of Transient response will be opened.
24. To identify parametric models, Polynomial Models is selected under Estimate popup menu. A dialog box appeared and there are some model structure provided in the "Structure:" to be generated. ARX: "[na nb nk]" is chosen and Estimate is pressed.

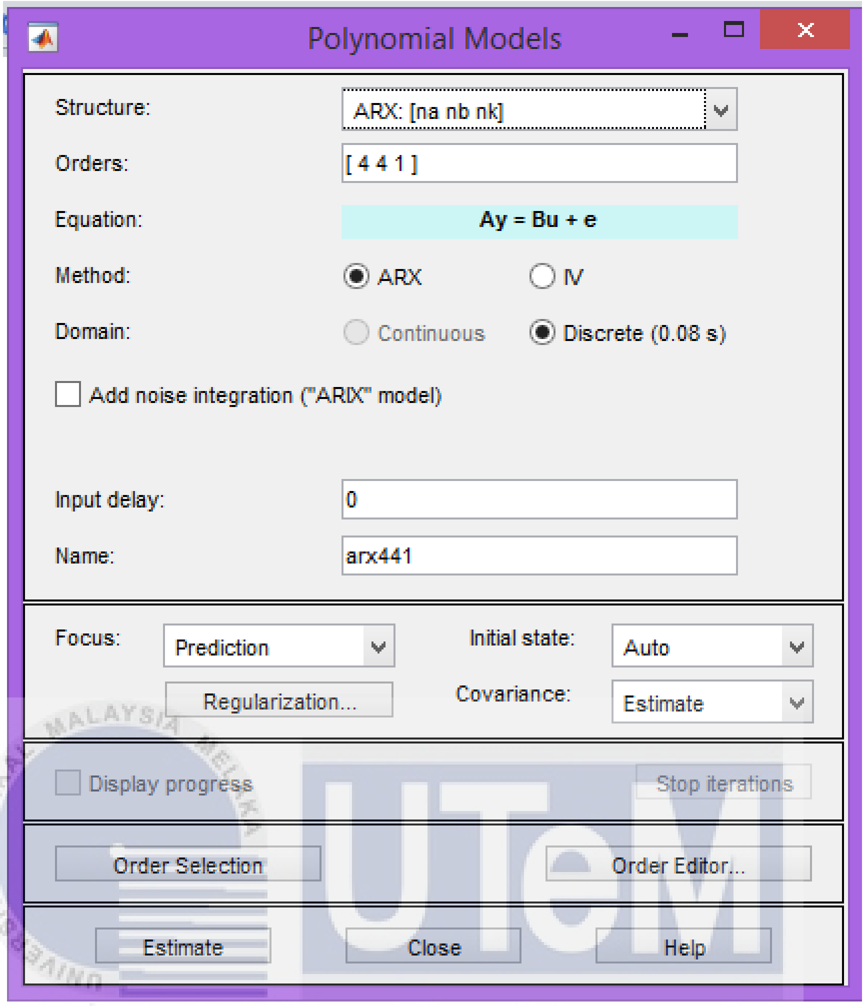


Figure 3.13: Polynomial Models

25. The model which is named arx441 is computed and appeared at the Model board as an icon.

Figure 3.14: Model board

26. At the Polynomial Models dialog box, the Order Editor is pressed to compute another model. The orders na=2, nb=2, and nk=3 in the Order Editor dialog box is selected. Another model, arx223 will be appeared at the Model board.

Figure 3.15: Model board

27. Transient resp at the Model Views is clicked and opened the Figure 3.16. There are three curves with colour coded along with the corresponding model icons in the Model board.

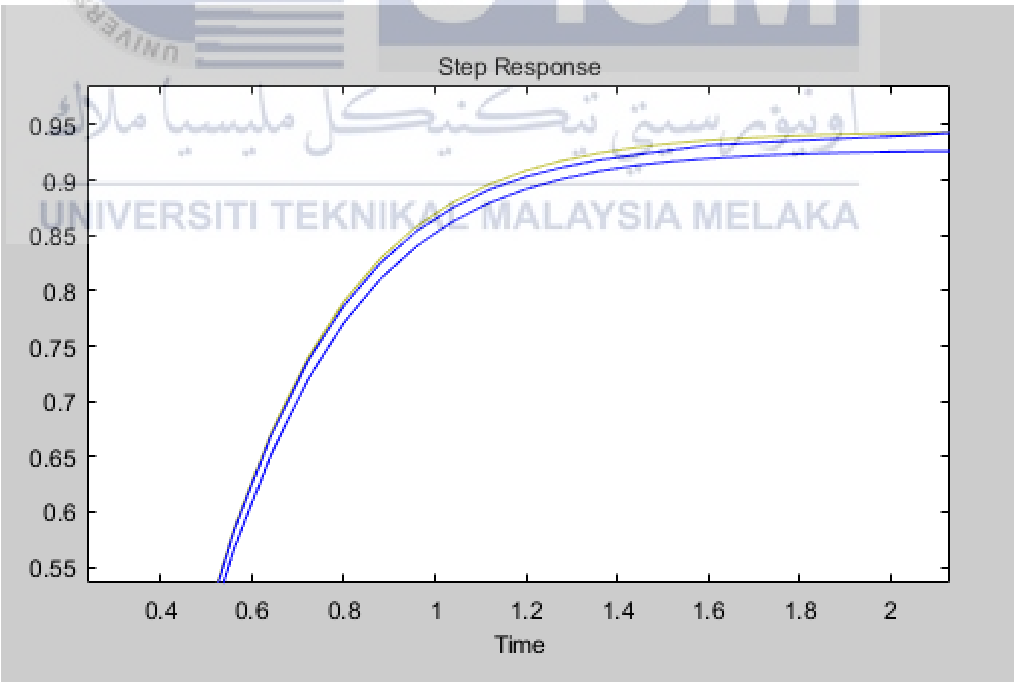


Figure 3.16: Transient response

28. Clicking on the model icons in the import model board, the associated curves are added or removed from the plot figure.

3.4 Performance Indicators

The objective of this study is to compare the modelling performance of ARX and Output-Error models based on selected indicators. Therefore, there are three indicators considered to determine the performance of model which are fit, Final Prediction Error (FPE), and Mean Square Error (MSE). The model that has the highest best fit is known as the best performance. Based on Akaike's theory, in a collection of different models, choose the one with the smallest FPE and MSE (Wikipedia, 2018). So, these indicators will be used in Chapter 4 for data analysis on the model performance. Figure 3.17 shows the example of best fits.

Figure 3.17: Best fit

CHAPTER 4

RESULT AND DISCUSSION

4.1 Introduction

This chapter will discuss the details about the findings that are obtained throughout this study. ARX model and Output- Error model have been studied in the simulation process done by using MATLAB software. The models are simulated by generating data for ARX model and Output-Error model by using the System Identification Toolbox.

4.2 Results, Analysis, and Discussion

4.2.1 Analysis using Elaborated Models for Data 1

Based on the trial run in Chapter 3, from the Polynomial Models dialog box, arx441 is computed by selecting the order na=4, nb=4 and nk=1 while another model is arx223 which is computed with na=2, nb=2 and nk=3. Both models' performance are compared based on selected indicators such as fit, Final Prediction Error (FPE) and Mean Square Error (MSE). The model that has a large value of best fits and small value of FPE and MSE are considered as a good performance model. For ARX type of model, different order and delay can be studied by the command arxstruc (Wikipedia, 2018). In this study, the first model specification is ARX352 as below.

$$y(t) = 0.2y(t-1) - 0.6y(t-3) + 0.5u(t-2) + u(t-6) + e(t)$$

Based on Akaike's theory, in a collection of different models, choose the one with the smallest FPE and MSE (Wikipedia, 2018). Figure 4.2 views models with different order and delay. As the result in Figure 4.3 and Table 4.1 show the one that has the highest best fits, smallest FPE and MSE have a good performance. In this first study of model, two different types of model which are ARX model and Output-Error Models are compared based on their best fits, FPE, and MSE. For example, in Table 4.1 shows that ARX352 has same value of best fits with OE532 and OE552 which is 98.25%. So, the other way to know which one that has the best performance is by looking at the value of FPE and MSE. Table 4.1 shows that OE532 and OE552 have greater value of FPE and MSE compared to ARX352. This shows that ARX352 has the best performance as it has the greatest fits and smallest in FPE and MSE values between the three models.

Structure	Fit (%)	Final Prediction Error (FPE)	Mean Square Error (MSE)
ARX352	98.25	0.0001412	0.0001341
ARX252	44.68	0.1693	0.1614
ARX342	9.745	0.2966	0.2839
OE512	44.72	0.1703	0.1663
OE532	98.25	0.0001744	0.0001656
OE552	98.25	0.0001757	0.0001654

Table 4.1: Data result of Fit, FPE, and MSE for Data 1 (Elaborated Models)

ARX252 and ARX342 has one less variable than ARX352 since ARX252 has na=2 while ARX352 is na=3. Then, OE512 has less variables than OE532 and OE552. OE512 has nf=1 while OE532 has nf=3 and this shows that OE512 has less variables there. Although OE552 has more variables, performance becomes stagnant where there are no more improvement. This means the value of nf order cannot be changed anymore but the value of nb and nk still can be changed to get a better performance.

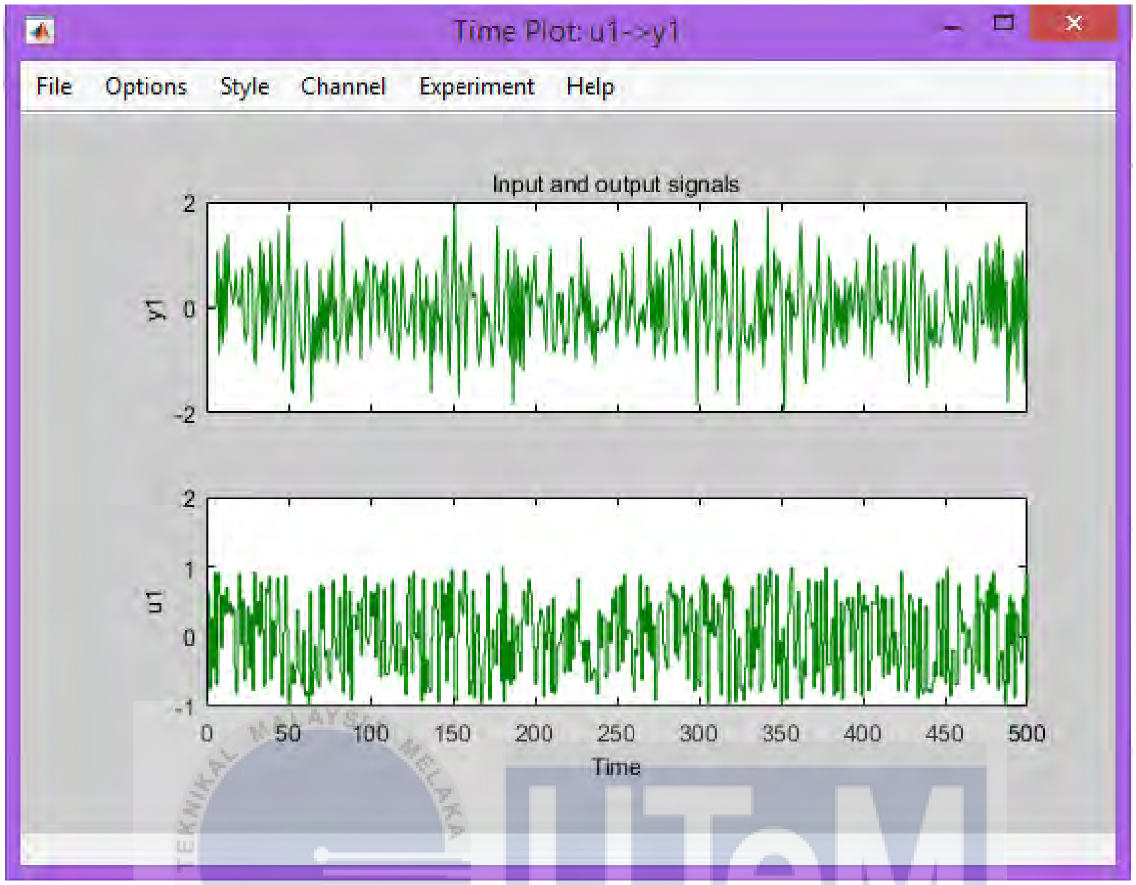


Figure 4.1: Time Plot for Data 1

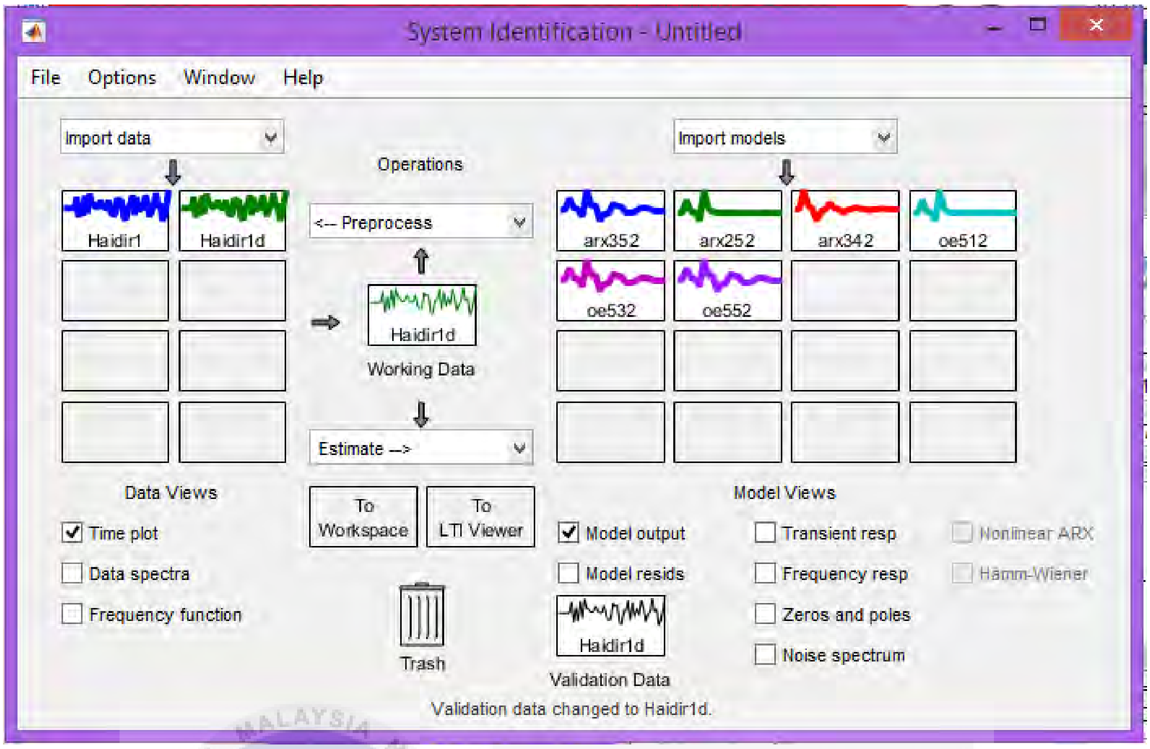


Figure 4.2: Models for Simulated Data 1

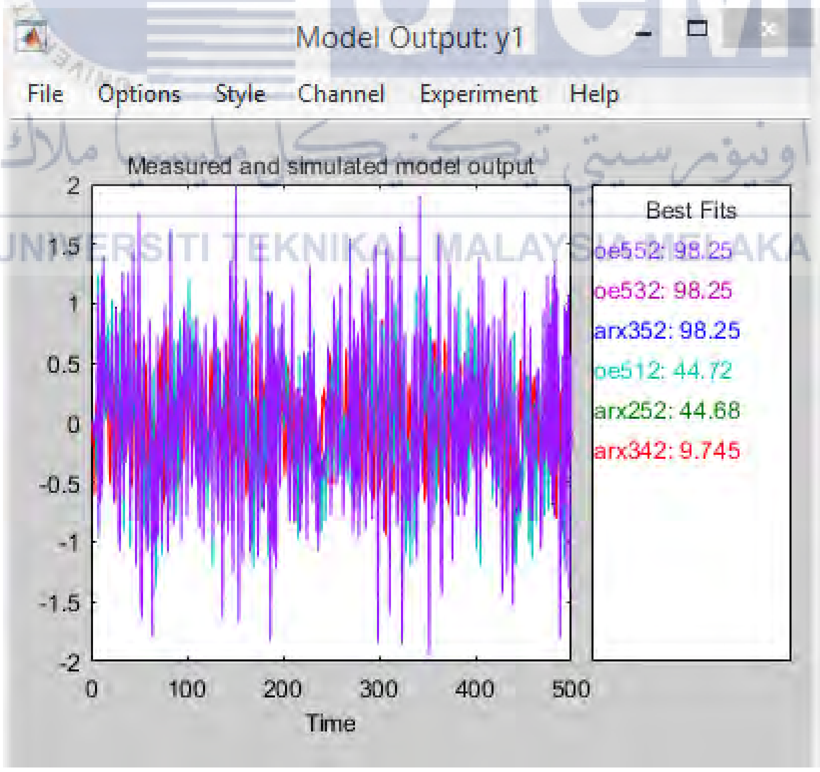


Figure 4.3: Best Fits for Data 1

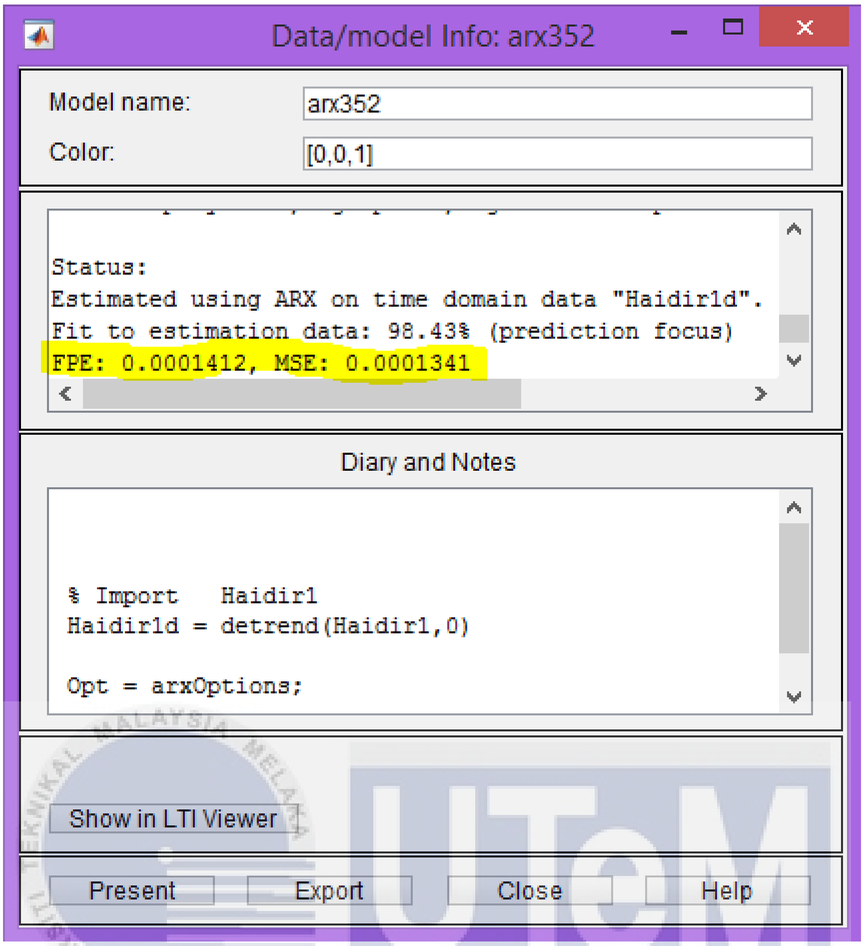


Figure 4.4: Data/Model info for Data 1

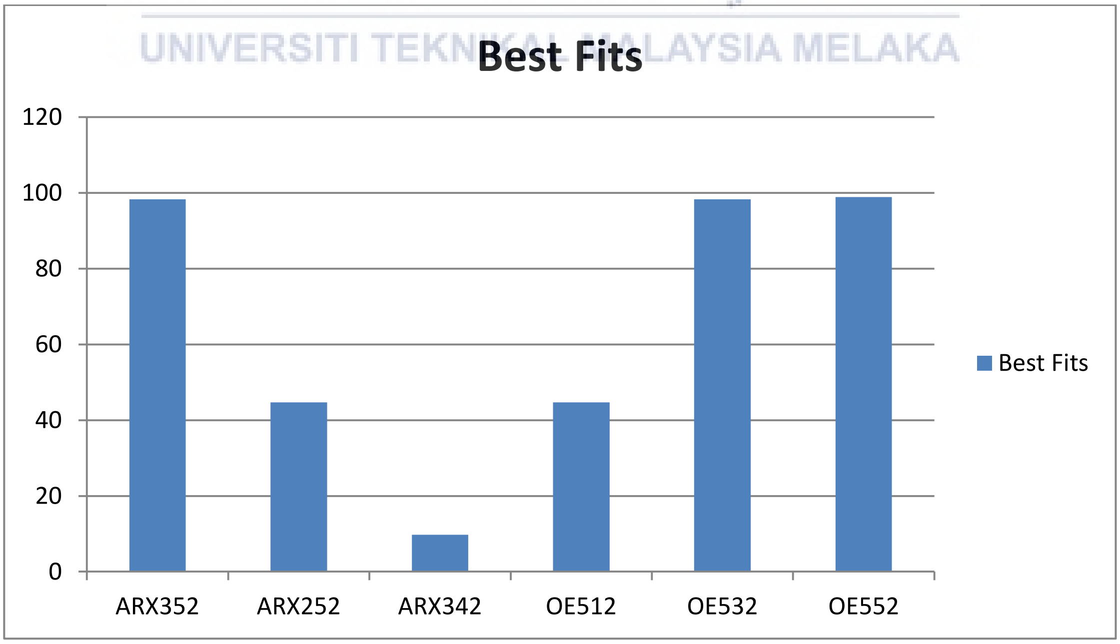


Figure 4.5: Best fits for Data 1

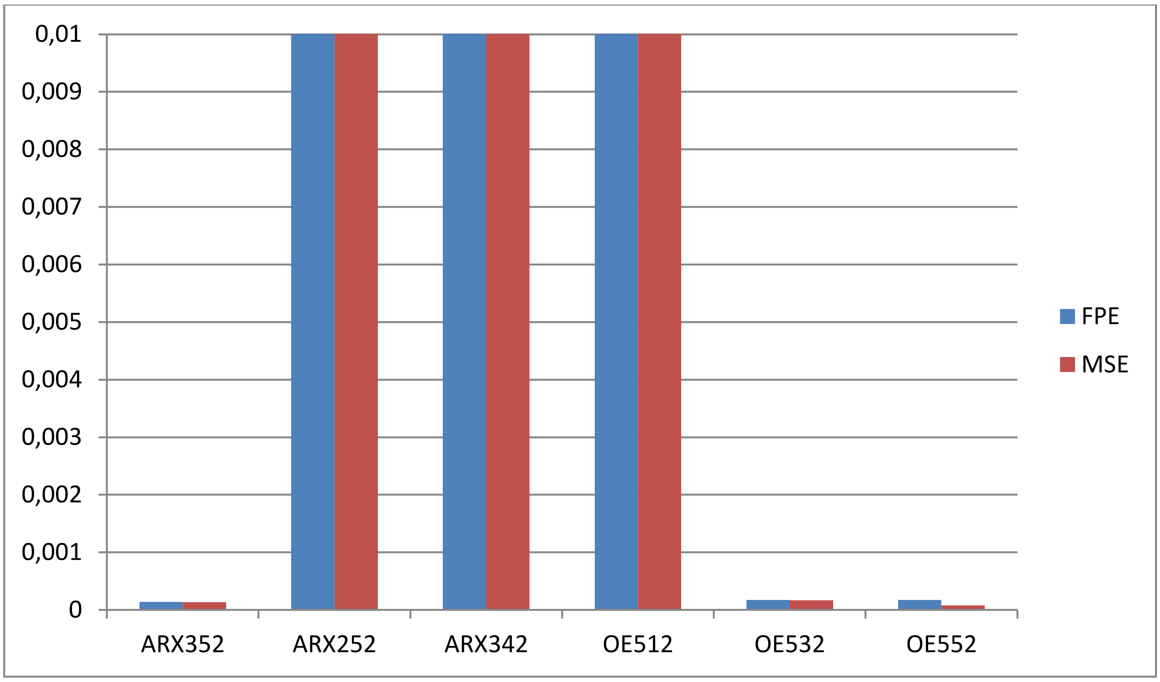


Figure 4.6: Final Prediction Error (FPE) and Mean Square Error (MSE) for Data 1

4.2.2 Analysis using Simple Models for Data 1

Figure 4.7 displays the time plot after preprocess by remove mean using another set of structures for the ARX model 1. In this section, ARX110, ARX210, ARX120, ARX130, OE110, OE210, OE120 are tried.

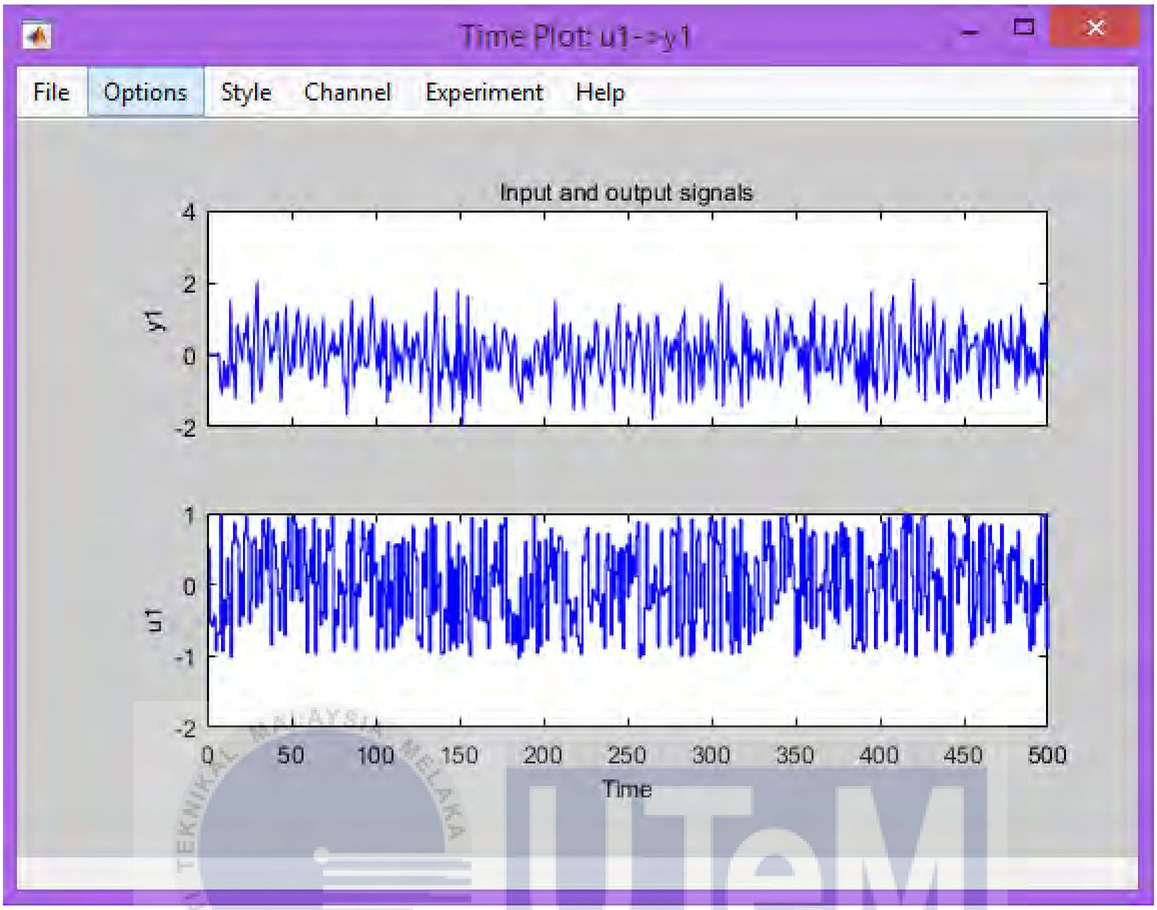


Figure 4.7: Time Plot for Data 1

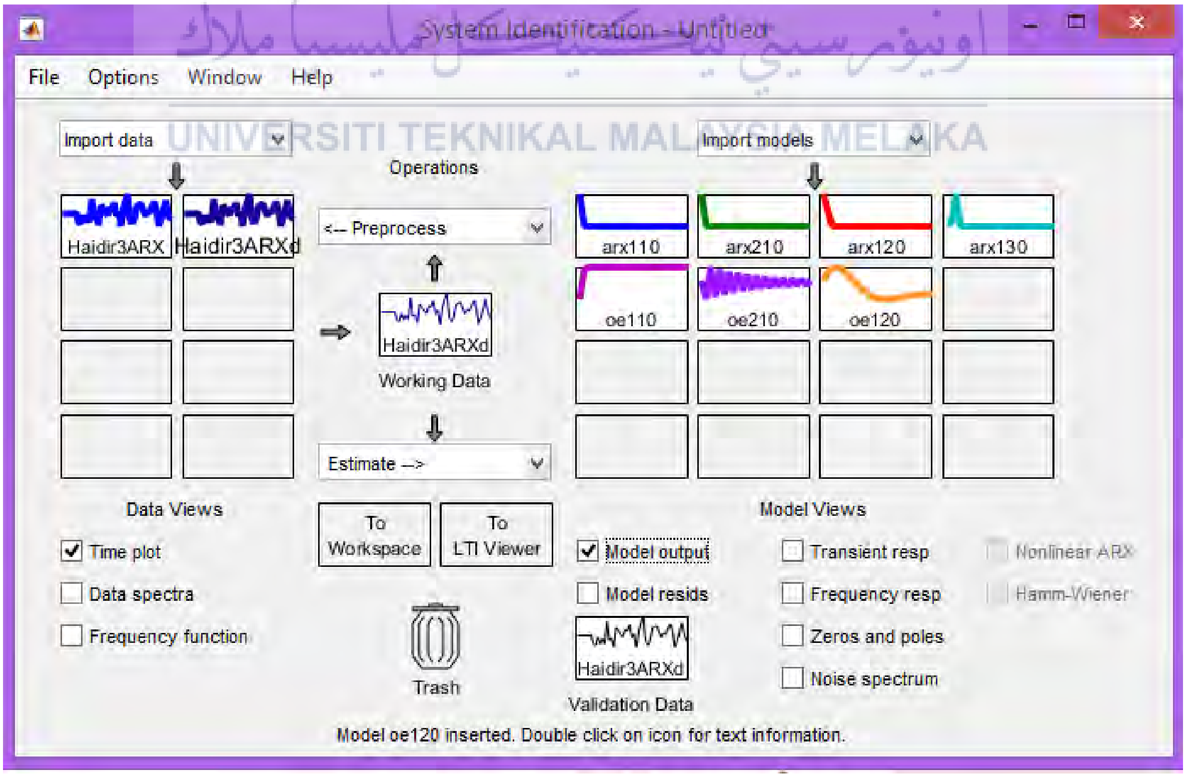


Figure 4.8: Simple Models for Simulated Data 1

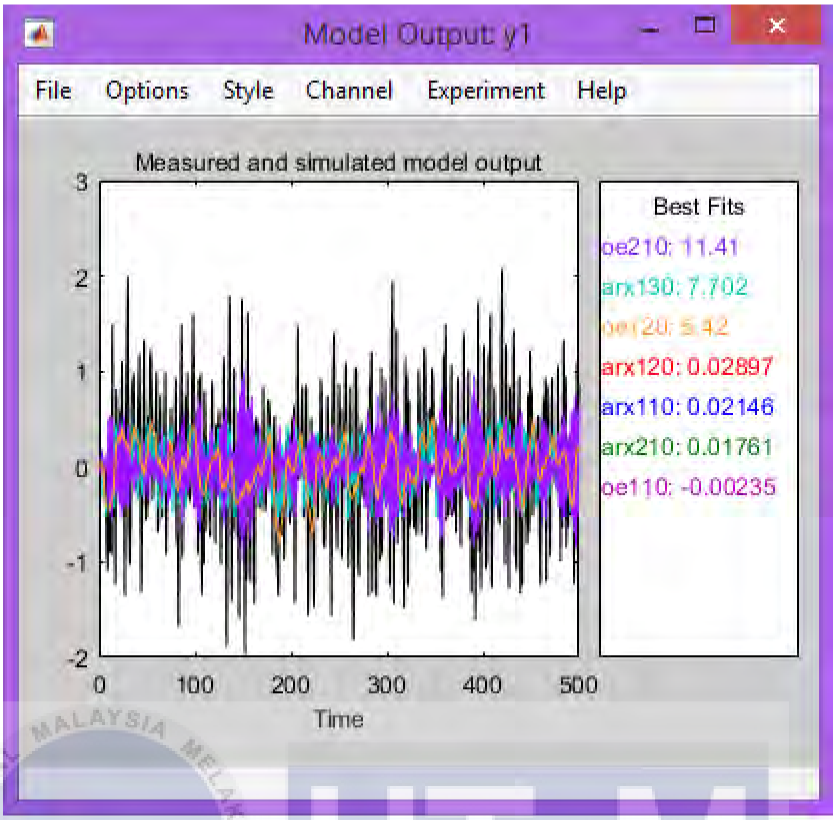


Figure 4.9: Best fits for Data 1

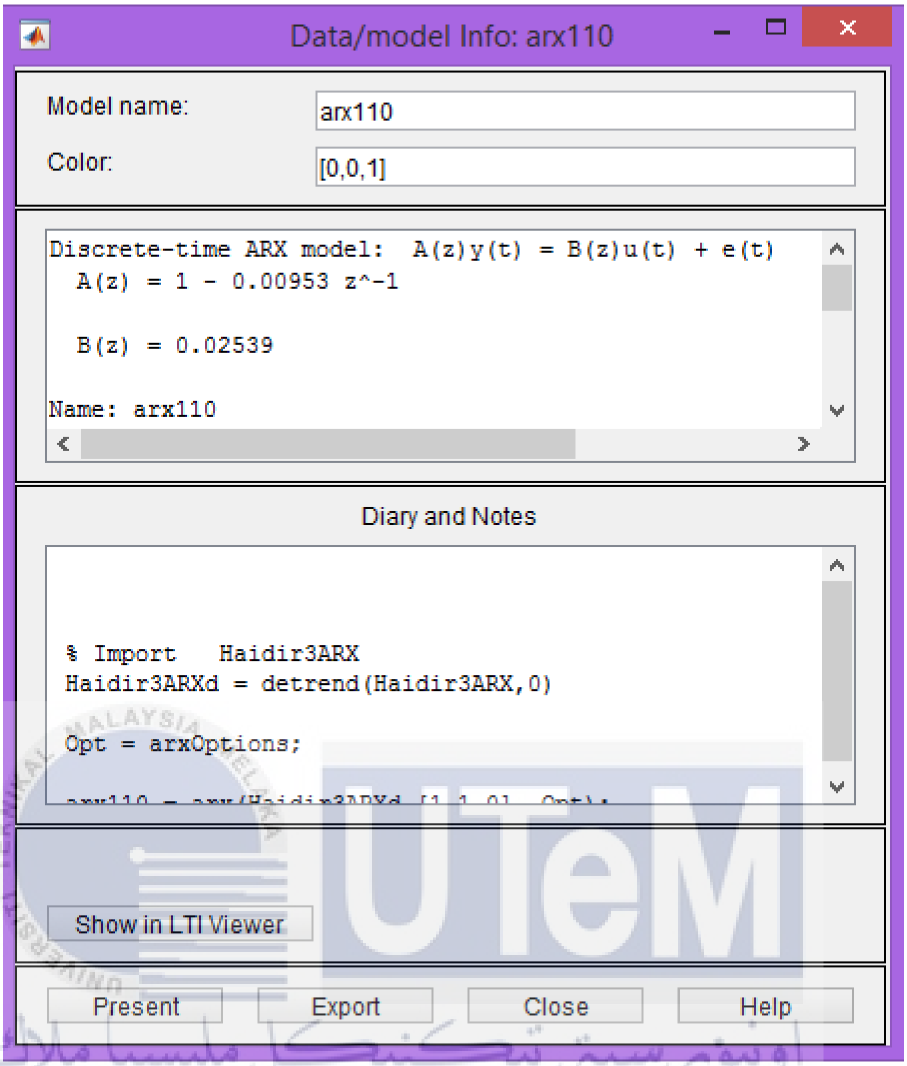


Figure 4.10: Data/Model Info for Data 1

Structure	Fits (%)	Final Prediction Error (FPE)	Mean Square Error (MSE)
ARX110	0.02146	0.5758	0.569
ARX210	0.01761	0.5805	0.569
ARX120	0.02897	0.5781	0.5689
ARX130	7.702	0.4969	0.4851
OE110	-0.00235	0.5739	0.5693
OE210	11.41	0.4523	0.4469
OE120	5.42	0.5156	0.5094

Table 4.2: Data result of Fit, FPE, and MSE for Data 1 (Simple Models)

Table 4.2 shows the data result obtained after generating data in the System Identification Toolbox as shown in Figure 4.8. By referring to Figure 4.9 can be seen that the highest best fits is OE210 which is 11.41%. This model also has smaller value of FPE and MSE. So, the best performance model is OE210.

ARX130 and OE210 have 4 variables and both of them are the highest fit among their own kind of structure. But, OE210 is better than ARX130 in term of performance due to its value of FPE and MSE. Then, OE110 and OE120 have one less variable than OE210. Meanwhile, ARX110 and ARX210 also have one less variable than ARX120.

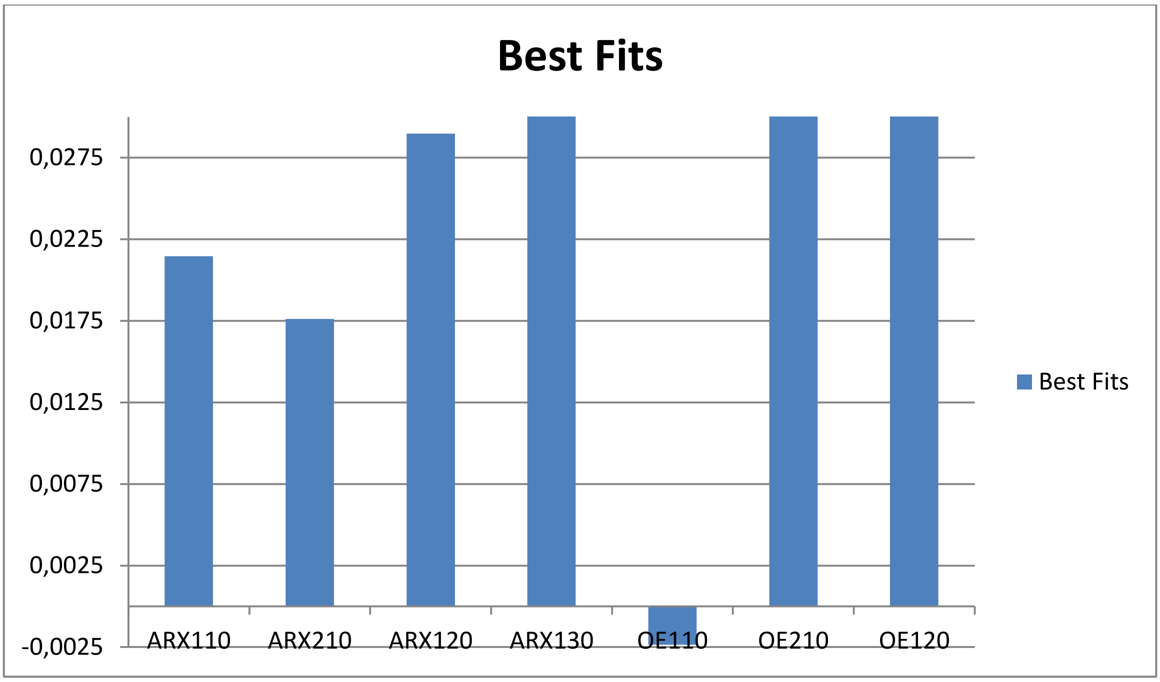


Figure 4.11: Best fits for Data 1

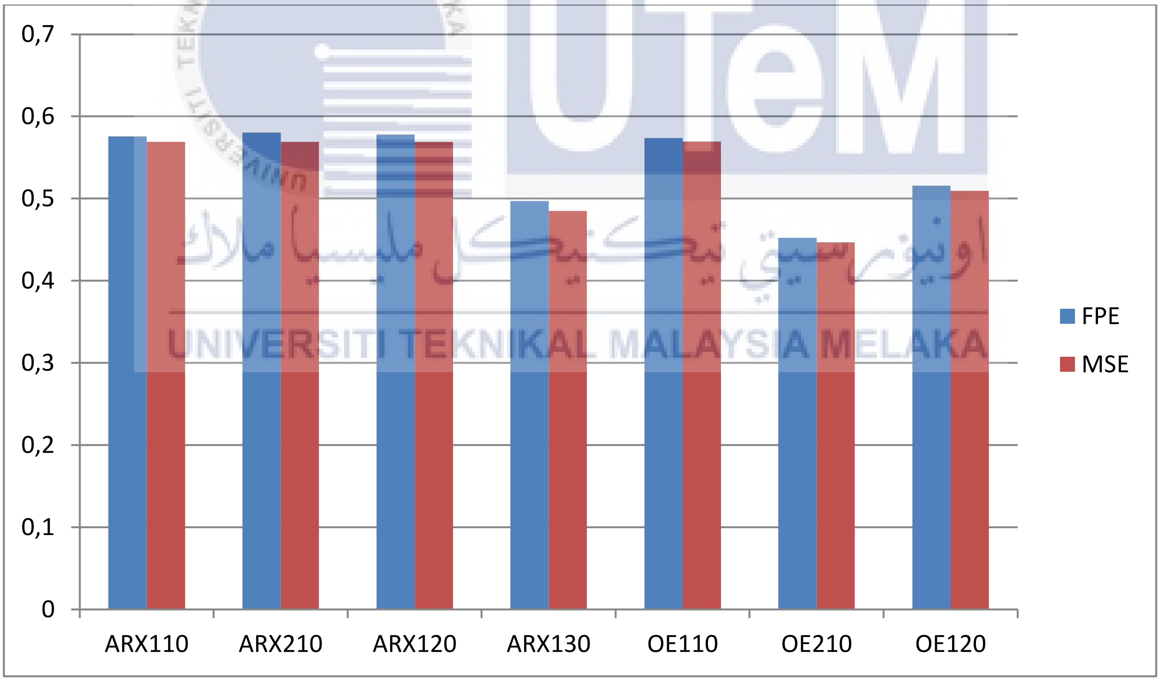


Figure 4.12: Final Prediction Error (FPE) and Mean Square Error (MSE) for Data 1

4.2.3 Analysis using Elaborated Models for Data 2

On other study investigates the performance for second model specification which is OE532 with its characteristic as below.

$$y(t) = \frac{0.5u(t-2) + u(t-6)}{0.3u(t) + 0.7u(t-1) - 0.1u(t-3)} + e(t)$$

Figure 4.13 shows the preprocess for a new data set at data board.

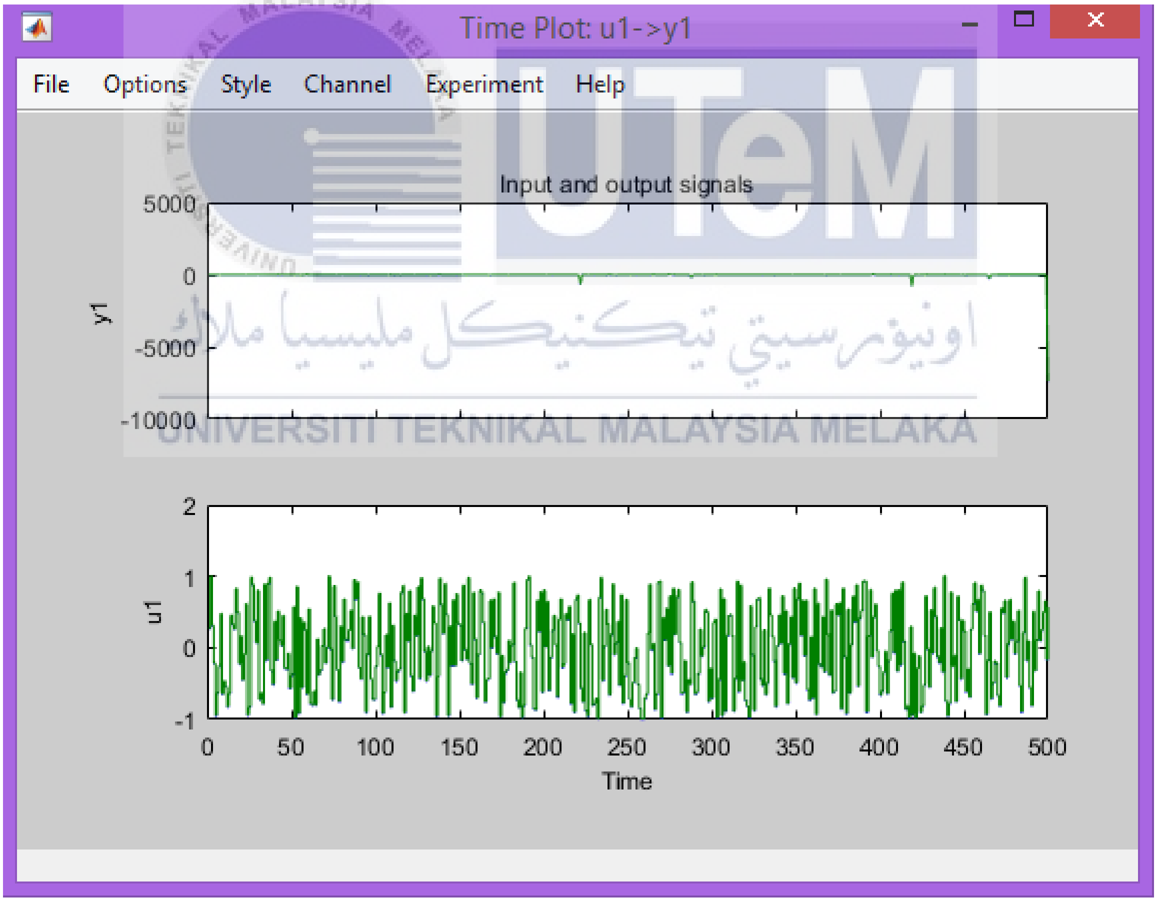


Figure 4.13: Time Plot for Data 2

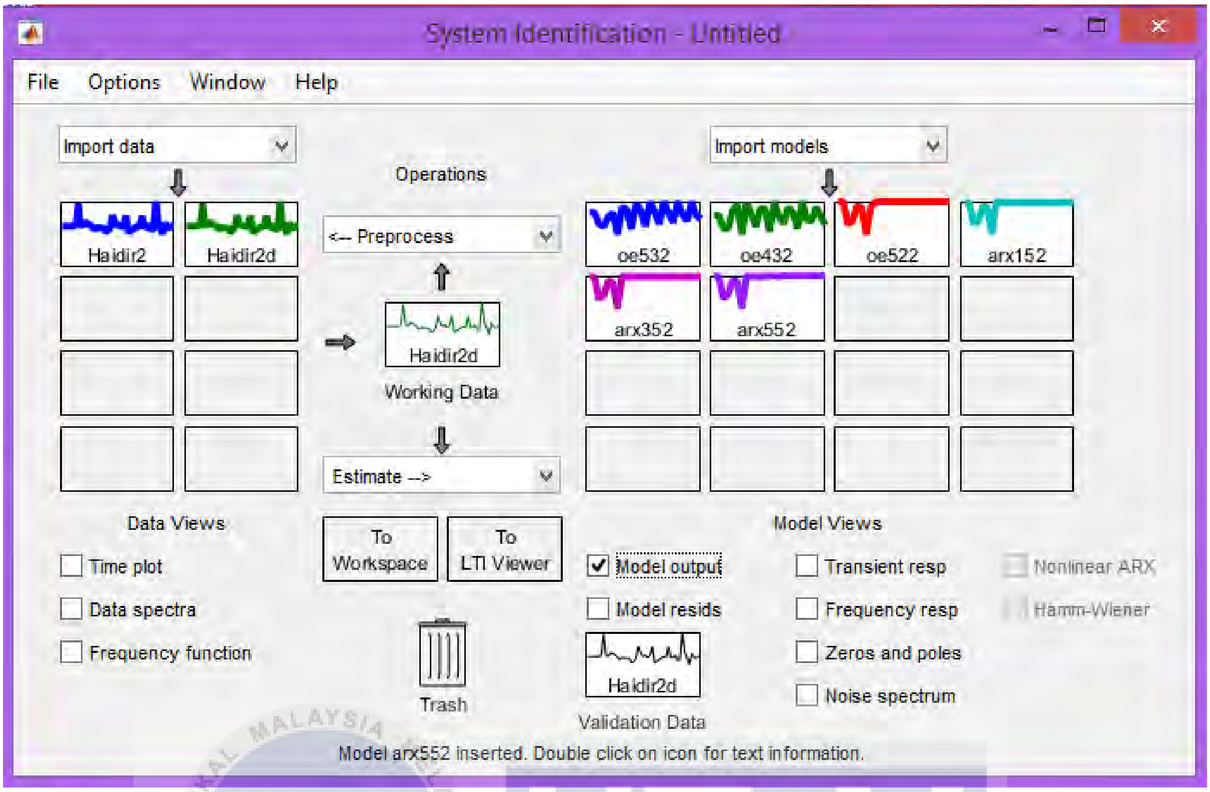


Figure 4.14: Models for Simulated Data 2

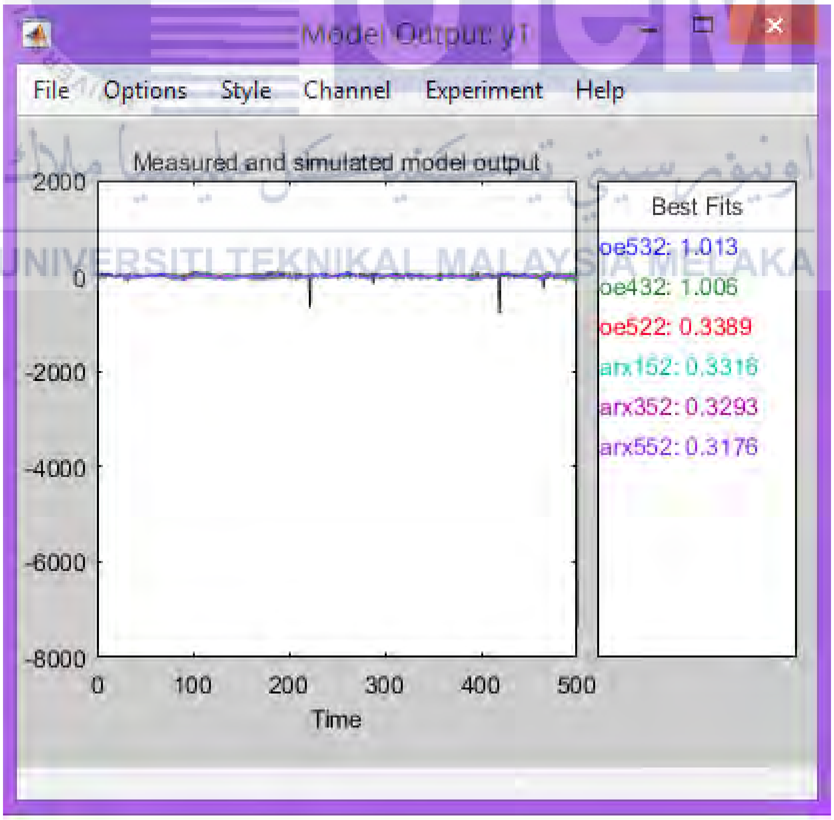


Figure 4.15: Best fits for Data 2

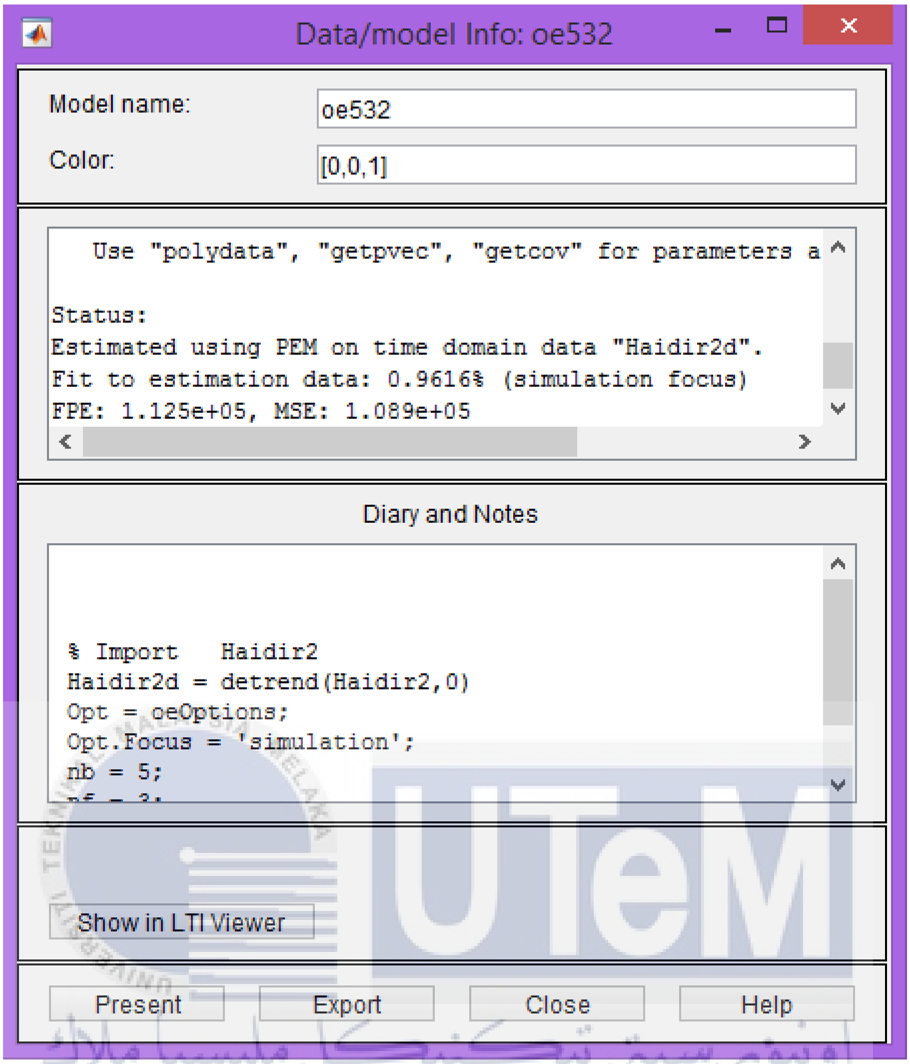


Figure 4.16: Data/Model info for Data 2

Structure	Fits (%)	Final Prediction Error (FPE)	Mean Square Error (MSE)
OE532	1.013	1.125e+05	1.089e+05
OE432	1.006	1.12e+05	1.089e+05
OE522	0.3389	1.134e+05	1.103e+05
ARX152	0.3316	1.153e+05	1.103e+05
ARX352	0.3293	1.162e+05	1.103e+05
ARX552	0.3176	1.171e+05	1.103e+05

Table 4.3: Data result of Fit, FPE, and MSE for Data 2 (Elaborated Models)

The data result obtained are shown in Table 4.3. In fact, the greater the fits value of a model then it has a better performance. But, back to Akaike's theory, in a collection of different models, choose the one with the smallest FPE and MSE (Wikipedia, 2018). Therefore, by looking the greatest value of fits in Table 4.3, OE532 has the same fits value with OE432 which is 1.0%. Figure 4.17 shows the bar chart of best fits for the OE532 and OE432. Besides, by referring to the value of FPE and MSE, the model that has the smallest value of FPE and MSE also are pointed to OE532 and OE432. Figure 4.18 displays the bar chart for models' FPE and MSE values. So, it can be concluded that OE532 and OE432 has the best performance.

ARX352 has one less variable than ARX552 since ARX352 has na=3 while ARX552 is na=5. Then, OE532 has more variables than OE432 and OE522. OE512 has nb=5 while OE432 has nb=4 and this shows that OE432 has less variables there. Meanwhile, OE522 has same na value with OE532 but has less value of nf=2 compare to OE532 which is nf=3.

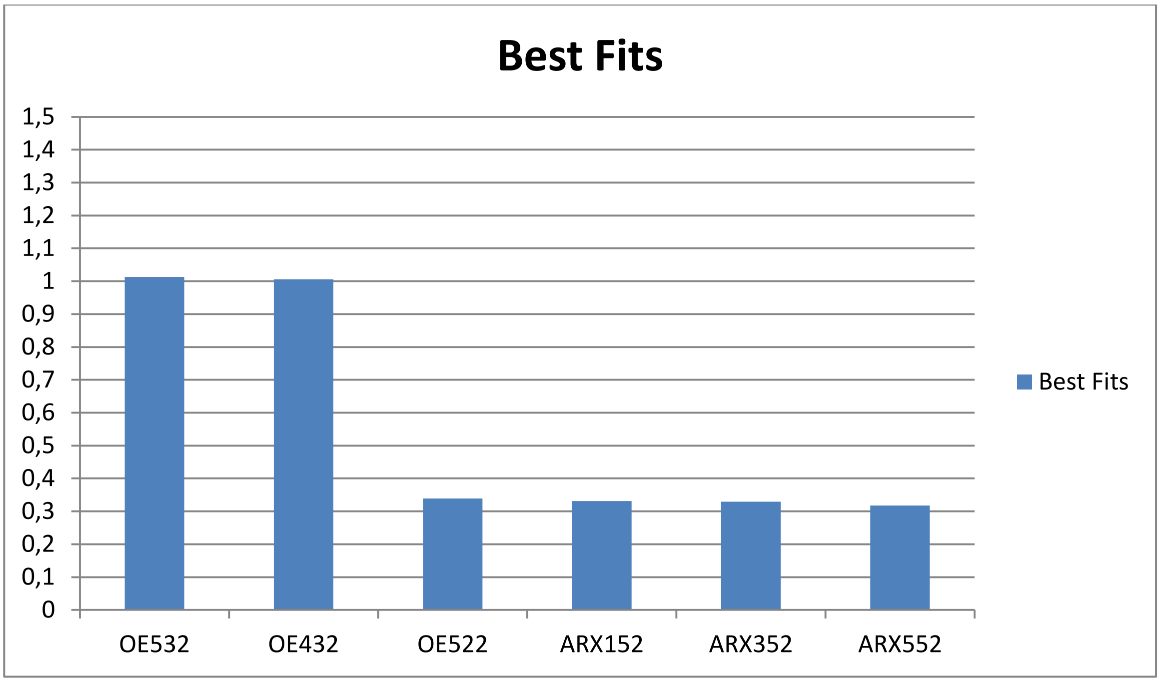


Figure 4.17: Best fits for Data 2

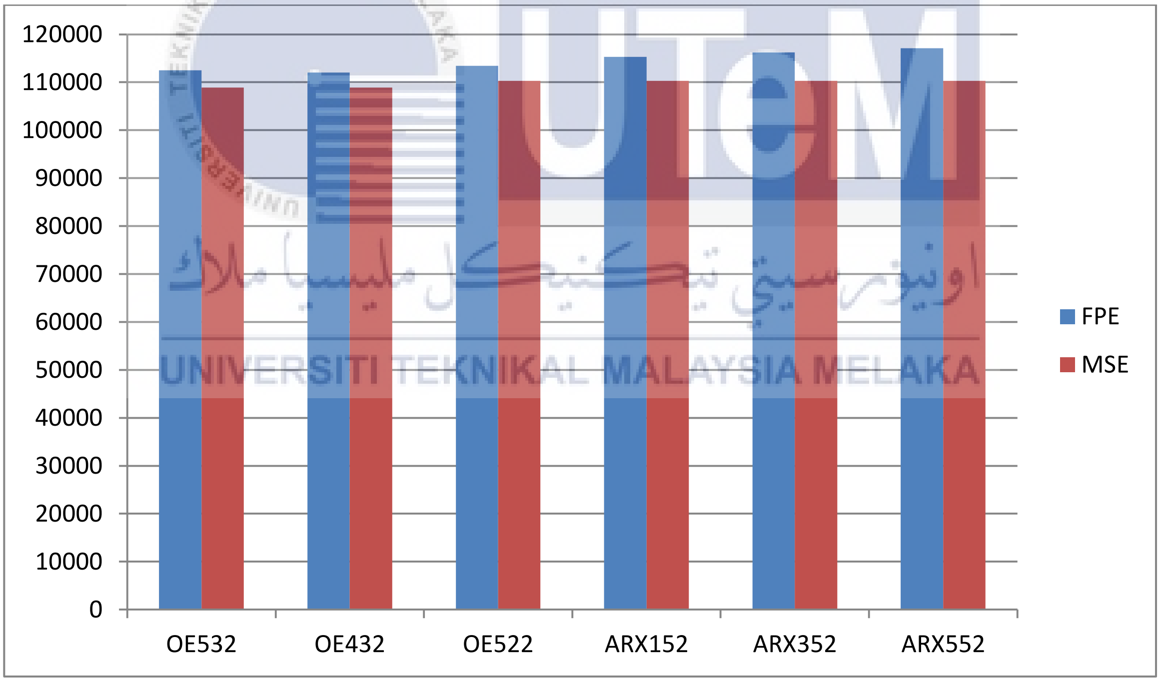


Figure 4.18: Final Prediction Error (FPE) and Mean Square Error (MSE) for Data 2

4.2.4 Analysis using Simple Models for Data 2

Figure 4.19 displays the time plot after preprocess by remove mean another set of structures for the Output-Error(OE) model structure. So in this section, ARX110, ARX210, ARX120, ARX130, OE110, OE210, OE120 are tried.

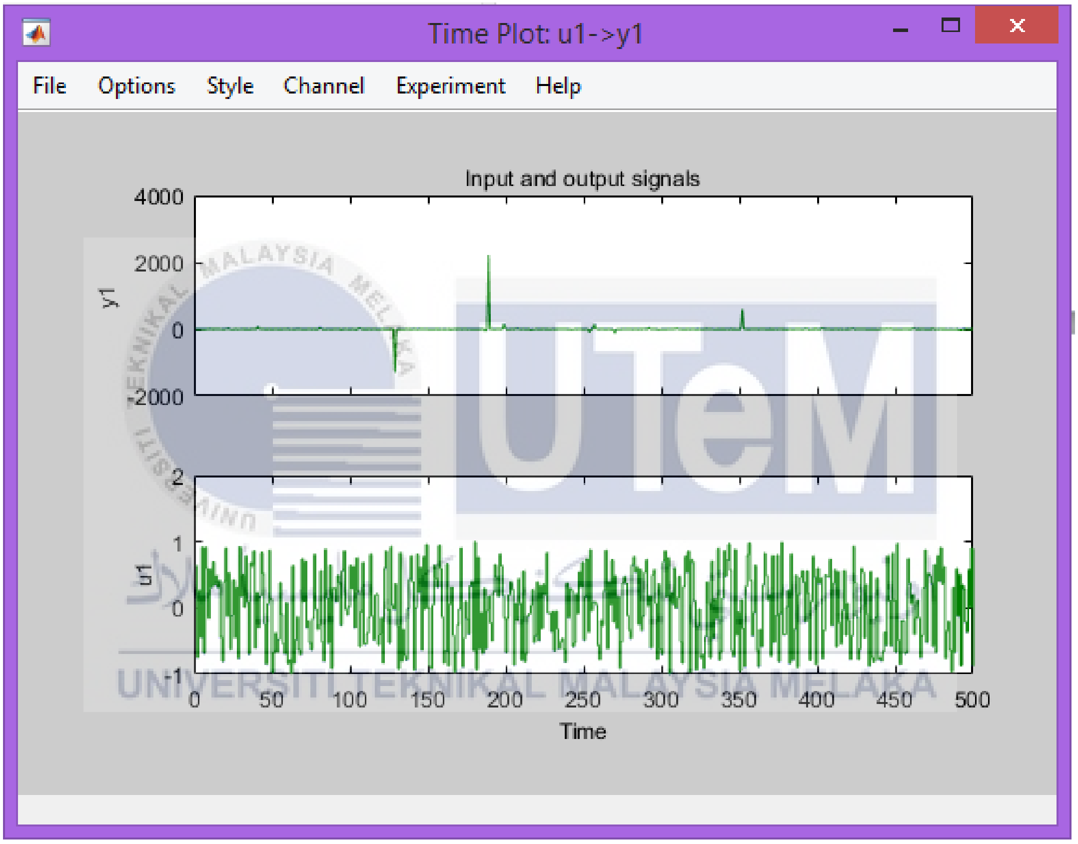


Figure 4.19: Time Plot for Data 2

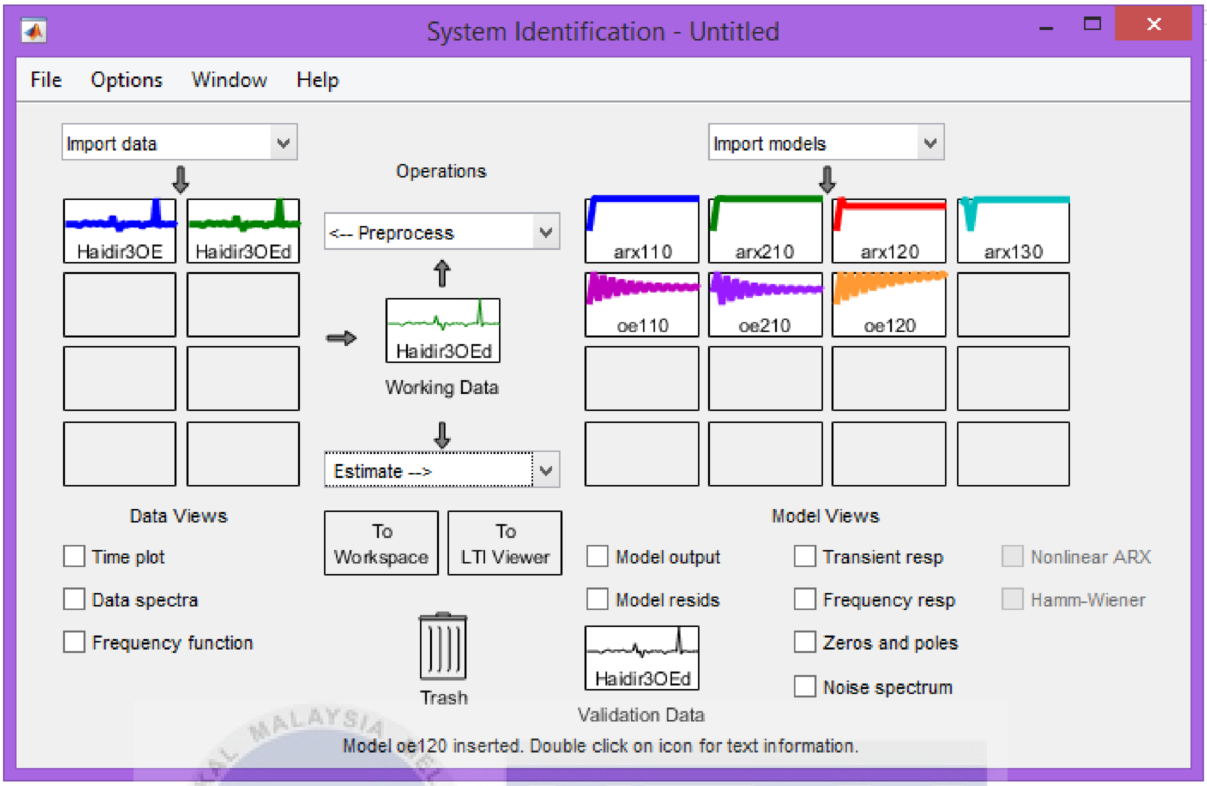


Figure 4.20: Simple Models for Simulated Data 2

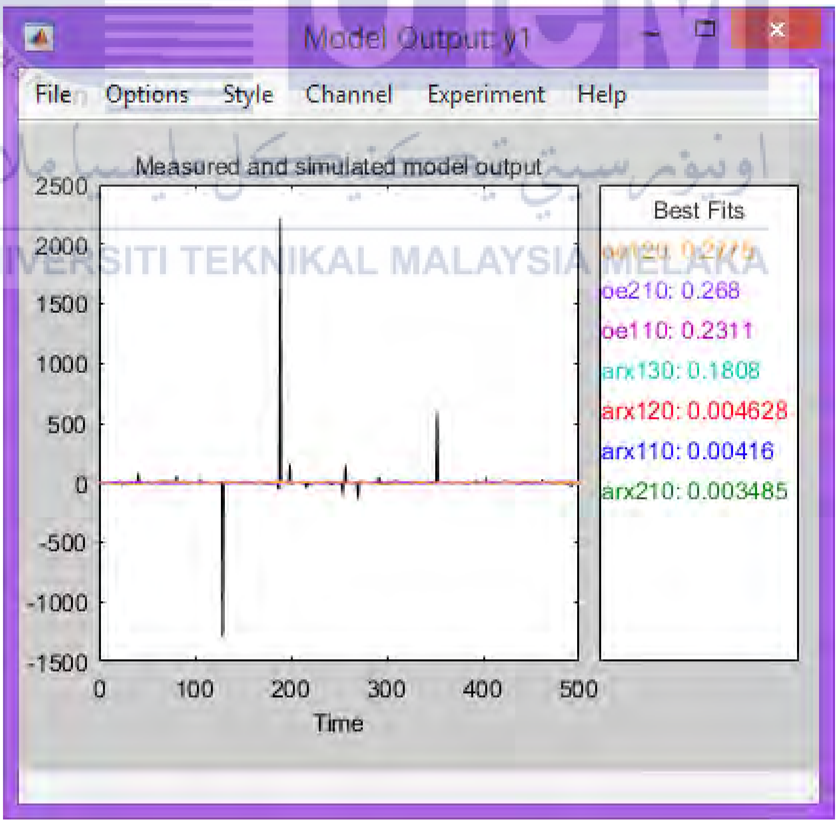


Figure 4.21: Best fits for Data 2

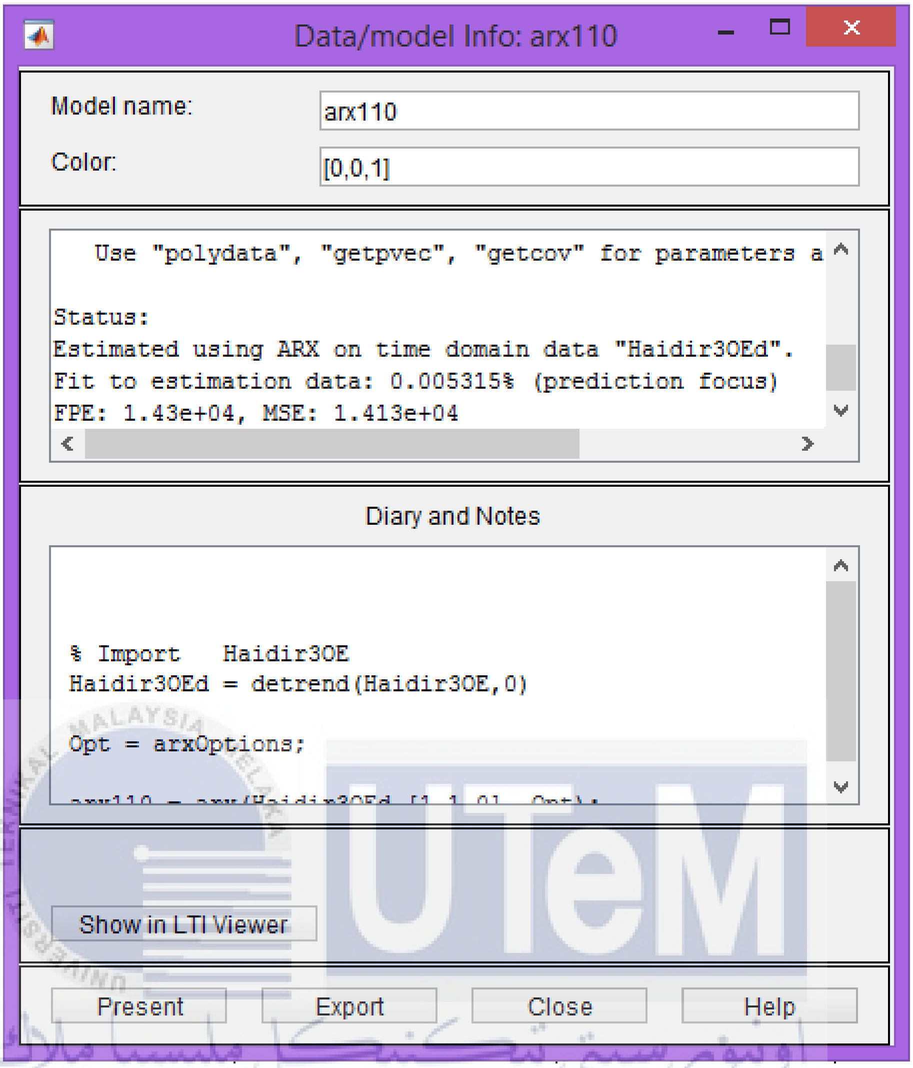


Figure 4.22: Data/Model Info for Data 2

Structure	Fits (%)	Final Prediction Error (FPE)	Mean Square Error (MSE)
ARX110	0.00416	1.43e+04	1.413e+04
ARX210	0.003485	1.441e+04	1.412e+04
ARX120	0.004628	1.435e+04	1.413e+04
ARX130	0.1808	1.442e+04	1.408e+04
OE110	0.2311	1.418e+04	1.406e+04
OE210	0.268	1.422e+04	1.405e+04
OE120	0.2775	1.422e+04	1.405e+04

Table 4.4: Data result of Fit, FPE, and MSE for Data 2 (Simple Models)

Table 4.4 shows the data result obtained in the System Identification Toolbox as shown in Figure 4.16. By referring to Figure 4.21, it can be seen that the highest best fits is OE120 which is 0.2775%. This model also has smaller value of FPE and MSE. Besides, OE120 and OE210 have same value of FPE and MSE which are 1.422e+04 and 1.405e+04 respectively. But, Figure 4.21 shows the highest fit is OE120. So, the best performance model is OE120.

ARX130 and OE120 are the highest fit among their own kind of structure because they have more variables. But, OE120 is better than ARX130 in term of performance due to its value of FPE and MSE. Then, OE110 and OE210 have one less variable than OE120. this is because nf order value for OE110 and OE210 is 1 compared to OE120 which is 2. Meanwhile, ARX110, ARX210, and ARX120 also have one less variable than ARX130. Most of the output data is near zero causing both types of models unable to capture the dynamics well.

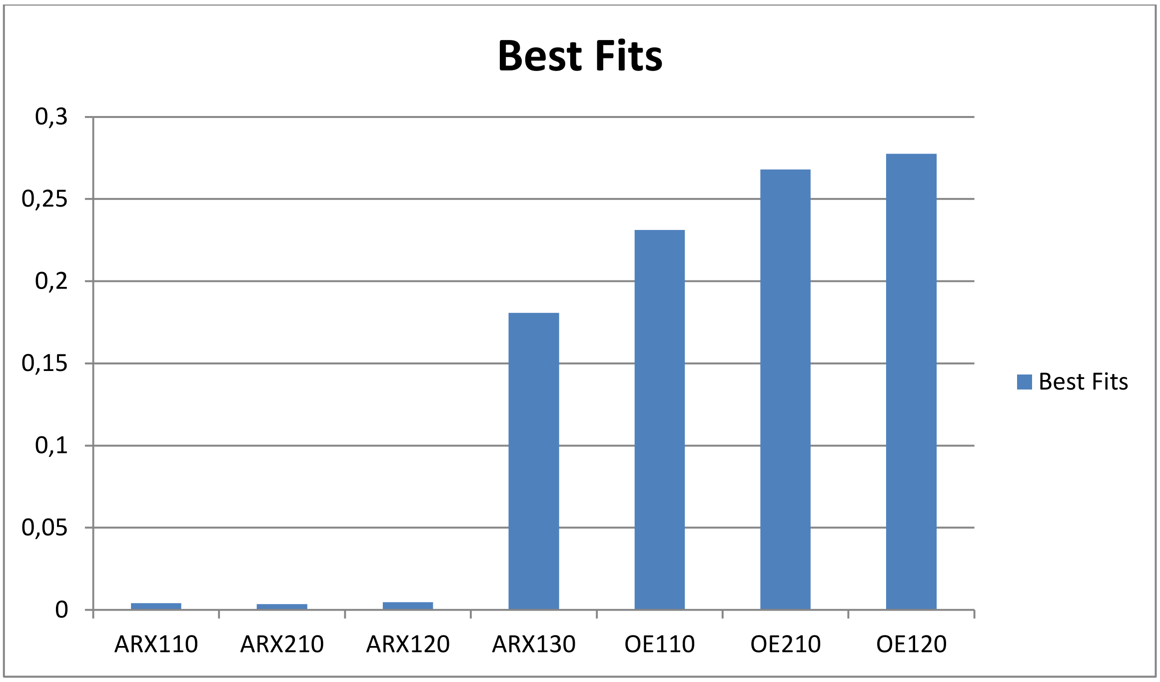


Figure 4.23: Best fits for Data 2

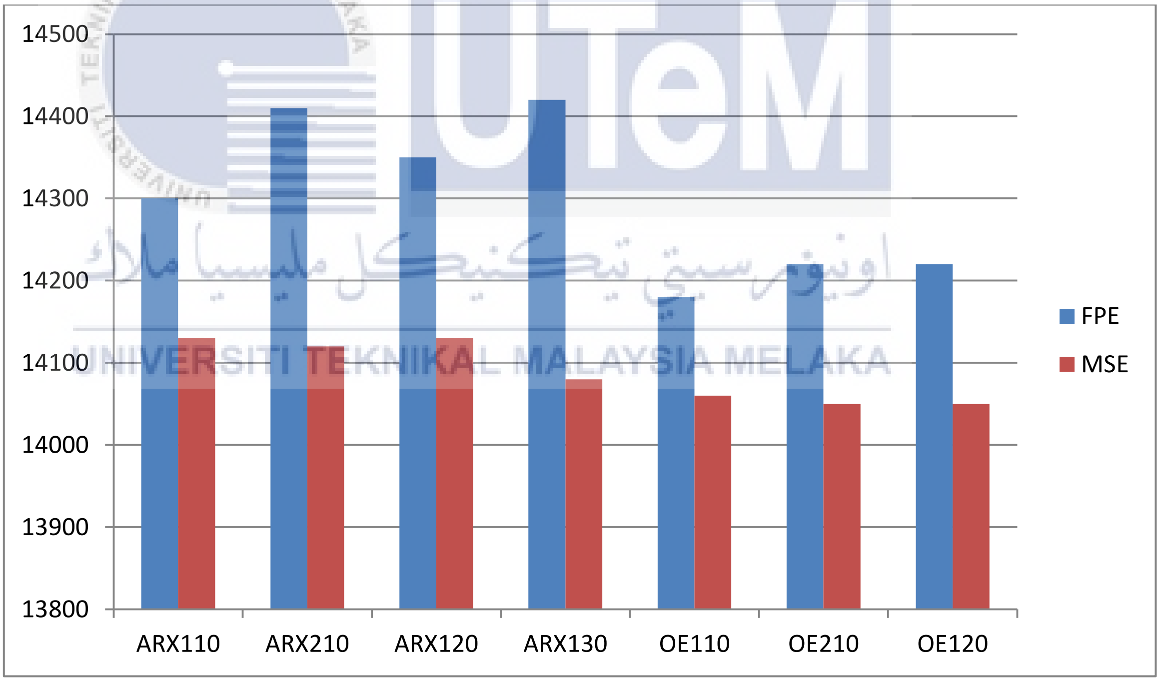


Figure 4.24: Final Prediction Error (FPE) and Mean Square Error (MSE) for Data 2

The bar chart in Figure 4.23 shows that the highest fit is OE120. Now, focus on each type of group, ARX130 is the highest fit among others ARX structure. Meanwhile, OE120 is the highest fit among others OE structure. Figure 4.24 displays the value of FPE and MSE. By only taking ARX130 and OE120, the FPE and MSE value for OE120 is the smallest.

4.3 Result of Real Data for ARX and Output-Error model

In this section studies about the performance that will be obtained when using a real data on dryer machine input and output data. So, the input, u data is the voltage of heating device and output, y data is output of air temperature. The sampling number used in this findings is 1000. In this study, three different data structure of ARX model and three data structure of OE model are tried to determine their own characteristics. They are ARX352, ARX252, ARX342, OE512, OE532, OE552. Figure 4.25 is the time plot graph for remove mean. There is input and output signals for the dryer.

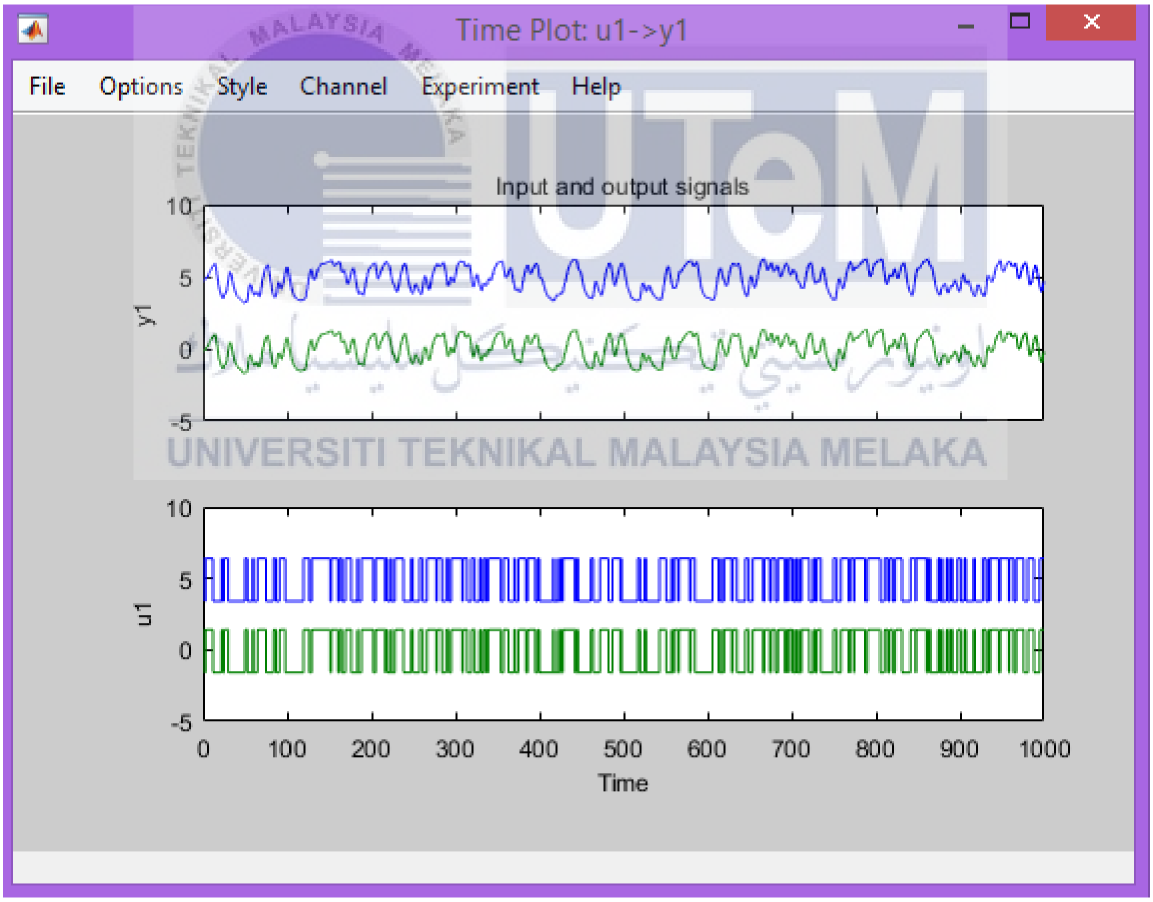


Figure 4.25: The time plot data (Remove Mean) for Real Data

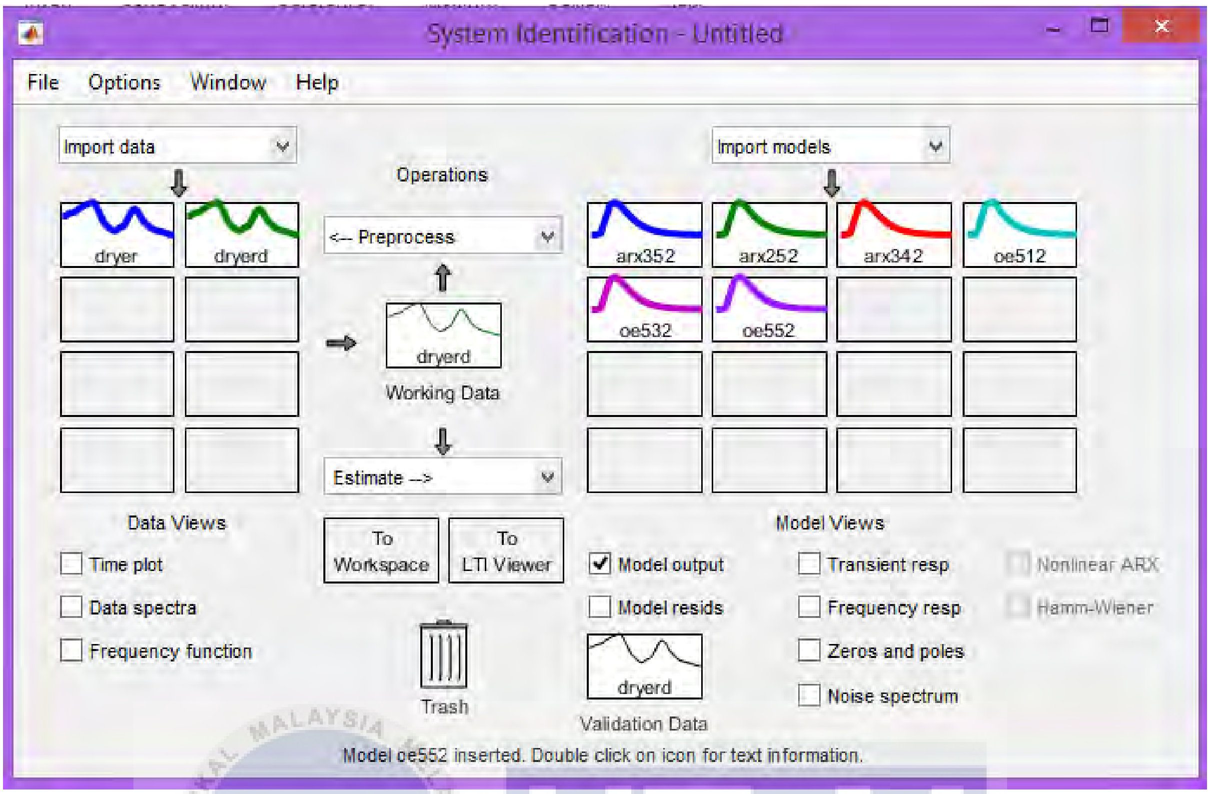


Figure 4.26: Models for Simulated Real Data

Figure 4.26 show all the six data that we obtain for this study. All this data show a same pattern in the model view board. Base on Figure 4.27, model structure OE552 have the highest fits value than other model structure. This is because if looking at the total variable number, the model structure OE552 have the highest variable number than other model structure. OE552 have six variable number while other model have five variable number.

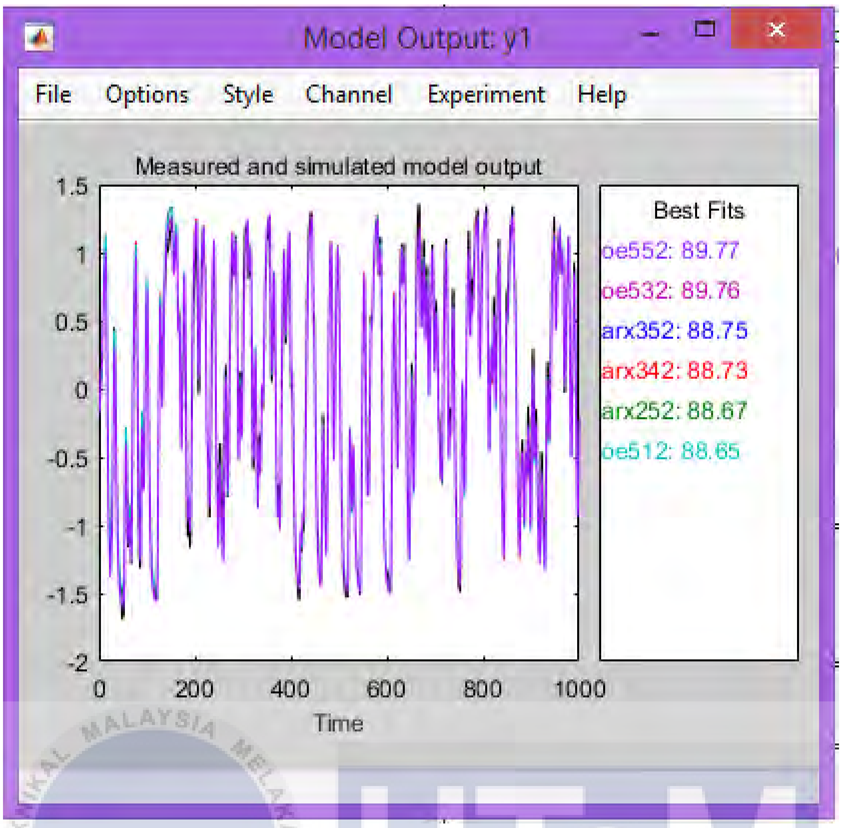


Figure 4.27: Best fits for Real Data

Structure	Fit (%)	Final Prediction Error (FPE)	Mean Square Error (MSE)
ARX352	88.75	0.00144	0.001403
ARX252	88.67	0.001482	0.001447
ARX342	88.73	0.00144	0.001409
OE512	88.65	0.009009	0.008901
OE532	89.76	0.008659	0.008522
OE552	89.77	0.008691	0.008519

Table 4.5: Data result of Fit, FPE, and MSE for Real Data

Table 4.5 shows the result for all the data value of fit, FPE, and MSE. From this result it shows that OE552 has the highest fit. By comparing OE552 with ARX352, the smallest value of FPE and MSE is in ARX352. But, OE552 has more variables than ARX352. Basically, more variables mean the larger the fit. Therefore, OE552 is the best performance model. Figure 4.28 and Figure 4.29 show the best fit, FPE, and MSE values.

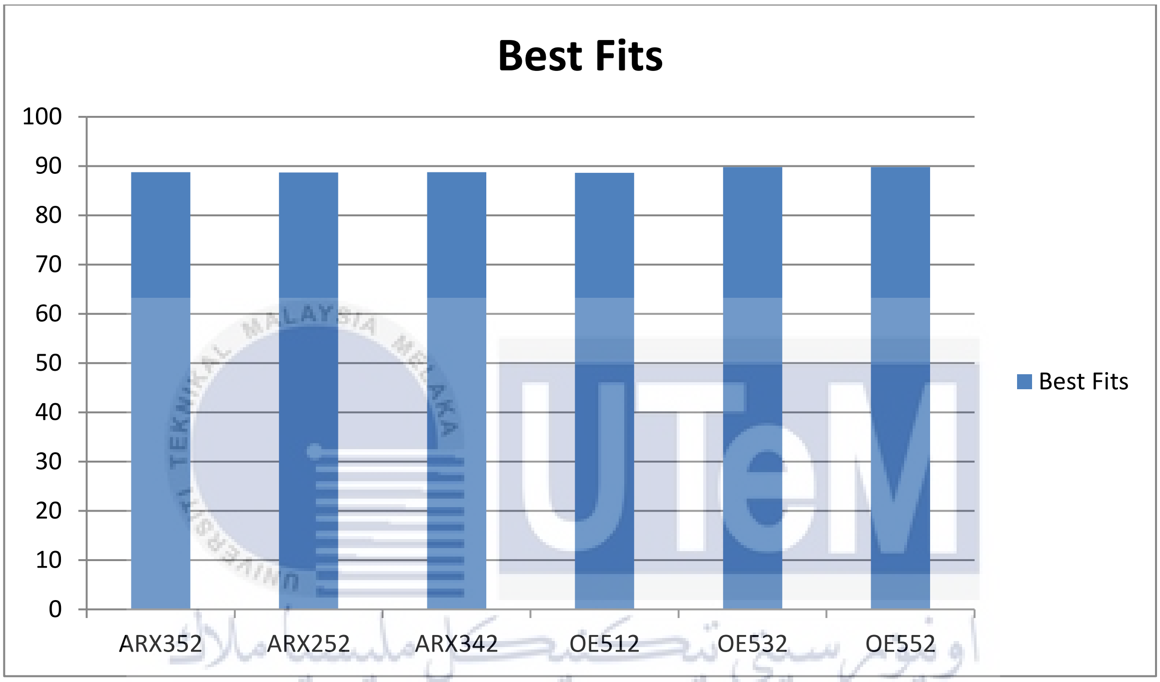


Figure 4.28: Best fits for Real Data

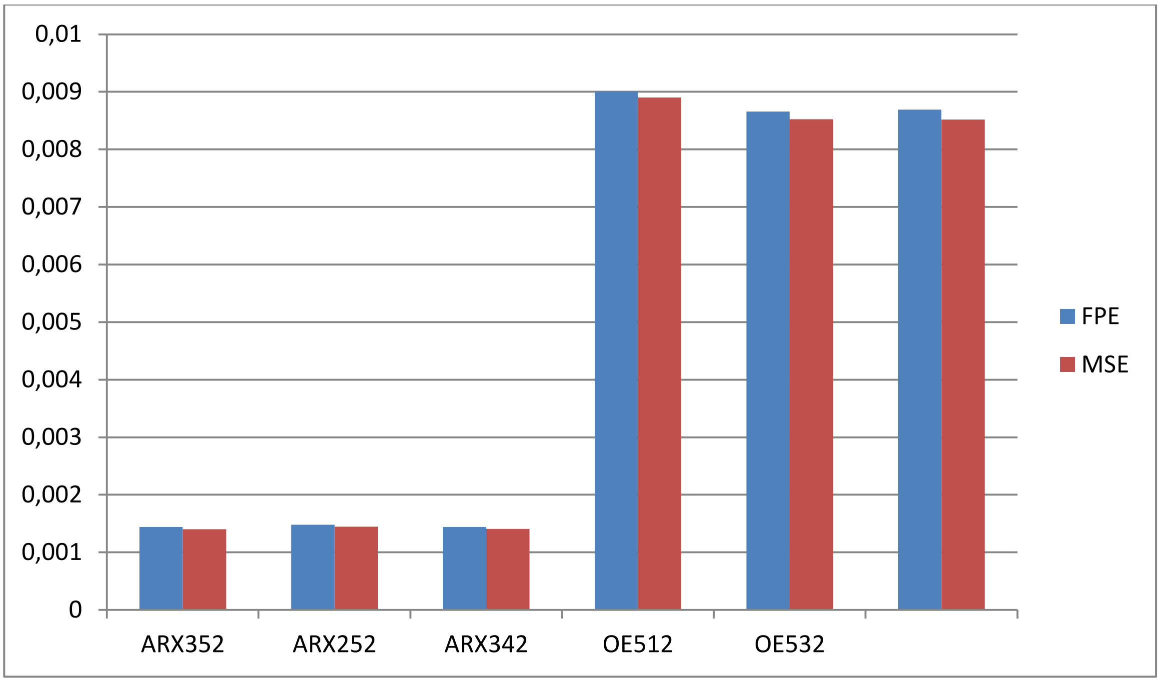


Figure 4.29: Final Prediction Error (FPE) and Mean Square Error (MSE) for Real Data

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study is about to simulate modeling using ARX and Output-Error models. The simulation is carried out by using MATLAB software as it has a function for System Identification. All the data have been generated in the System Identification Toolbox. The simulation is done to compare the modeling performance of ARX and Output-Error models based on the selected indicators. The selected indicators are fit, Final Prediction Error(FPE) and Mean Square Error(MSE). The performance of model structure can be determined by referring those three indicators. The analysis is done to gain the behavior of model with different input and output data. From the result show, the main factor that can affect the fit value is total variable number. It show that the higher value for total variable number, the better the performance of the modelling. Based on two different specific data and also one real data from industry, it show that every model structure that have highest variable number will get the highest fits value and lowest value for MSE and FPE.

However, most of the fit is near zero causing both types of models unable to capture the dynamics well. The best modelling performance is the model that have the highest fits value.

5.2 Recommendation

MATLAB software is highly recommended in this study as it provides the System Identification function. The simulation also need to be done with the same method in order to get a reliable result. The fit value will be small number if the input and output data are not suitable with the model structure. The right selection of model structure will get highest fit value nearest to 100 percentage fits value and has the smallest value of FPE and MSE. Besides, deeper study needs to be carried out in order to obtain a suitable model for data. Other than that, user can try another estimator other than least square method.

REFERENCE

str m, K. ., ittenmark, B. 1 . *Computer controlled systems theory and design: solutions manual*. Upper Saddle River (N.J.): Prentice-Hall.

Ding, F., Shi, Y., & Chen, T. (2007). Auxiliary model-based least-squares identification methods for Hammerstein output-error systems. *Systems & Control Letters, 56*(5), 373-380. doi:10.1016/j.sysconle.2006.10.026

Franklin, G. F., Powell, J. D., & Workman, M. L. (1998). Digital control of dynamic systems. *World Student Series*, 1–12. https://doi.org/10.1109/TASSP.1981.1163502

Ljung, L. (2010). System Identification Toolbox ™ Use r ' s Guide. *Order A Journal On The Theory Of Ordered Sets And Its Applications*, 566.

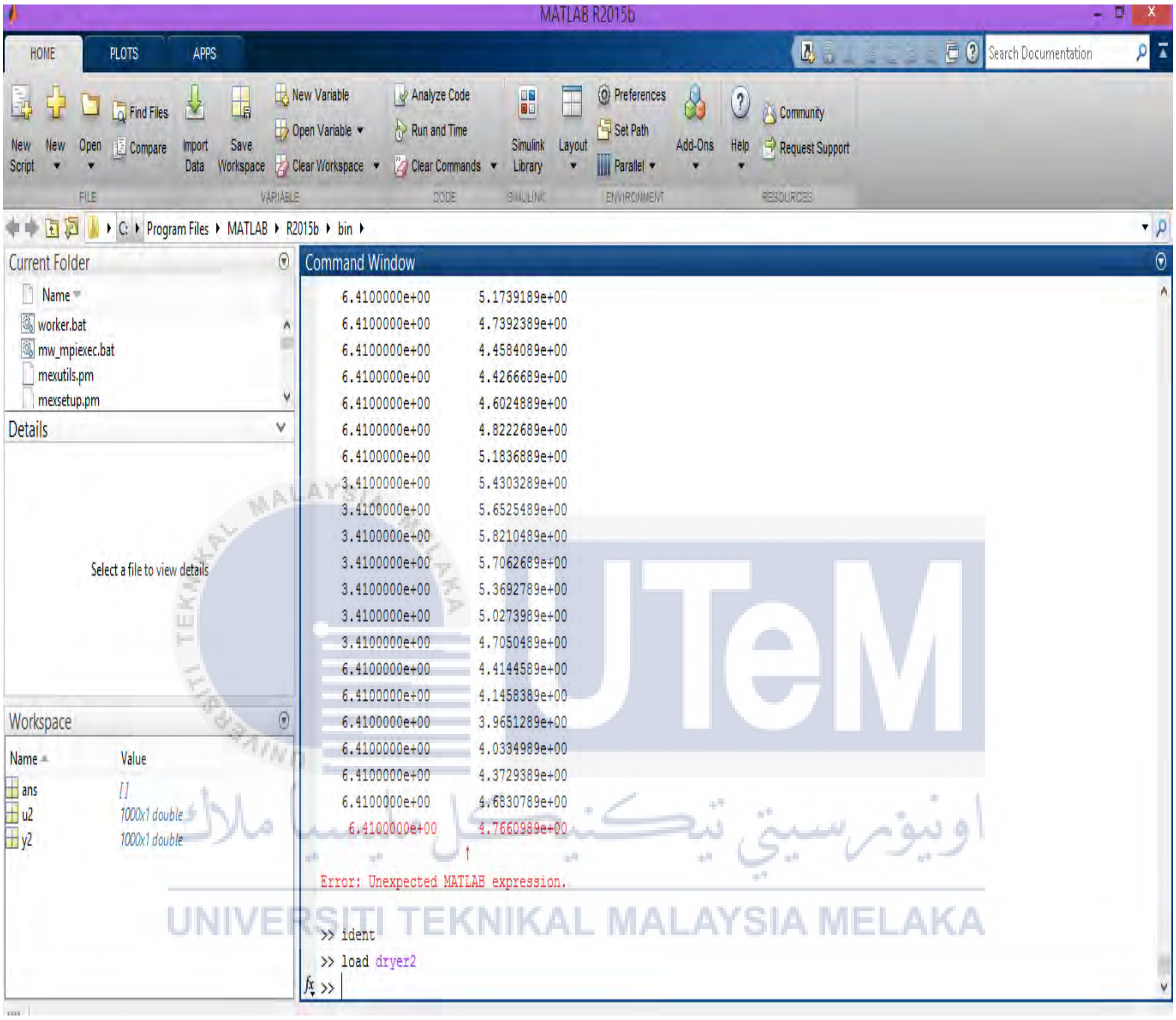
(n.d.). Retrieved March 22, 2018, from http://radio.feld.cvut.cz/matlab/toolbox/ident/ch3tut36.html

Pintelon, R., & Schoukens, J. (2012). *System identification a frequency domain approach*. Hoboken, NJ: John Wiley & Sons Inc.

Staff, I. (2016, May 17). Least Squares Method. Retrieved from https://www.investopedia.com/terms/l/least-squares-method.asp

APPENDICES

A) Real data inserted in command window



B) GANTT CHART PSM 1

WEEK / ACTIVITIES	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Research Title Selection	■	■				MIDTERM BREAK								
Literature Review	■	■	■	■	■		■	■	■	■	■	■	■	■
Familiarization with MATLAB Environment				■	■									
Submission of Progress Report 1							■							
Trial Run By Using MATLAB								■	■	■				
Discussion on Trial Run Result									■	■	■			
Report writing											■	■		
Report Submission													■	
PSM 1 Seminar														■

C) GANTT CHART PSM 2

WEEK / ACTIVITIES	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Data Acquisition															
Literature Review															
Simulation of System identification															
Submission of Progress Report 1															
Result analysis						MIDTERM BREAK									
Analysis and Discussion															
Conclusion															
Report writing															
Report Submission															
PSM 2 Seminar															